

[illegible]

DR N-PSDB; AAC85697.
XX
XX New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT substrates, identifying inhibitors of DPP8 catalytic activity which
PT have therapeutic uses, and for detecting activated T cells -
XX
XX Claim 5; Page 74-75; 78pp; English.
XX
XX The sequences given in AAB47187-90 represent fragments of human
CC dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for
CC H-Ala-Pro-pNA, H-Gly-Pro-pNA and H-Arg-Pro-pNA. Therefore, it is a
CC prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable
CC of hydrolysing the peptide bond C-terminal to Pro in each of these
CC compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for
CC cleaving a substrate, and for detecting an activated T cell which
CC involves measuring the level of DPP8 gene expression in a T cell. The
CC level of DPP8 expression is detected by detecting the amount of DPP8
CC RNA in the cell. It is also useful for identifying a molecule capable
CC of inhibiting the cleavage of the substrate by DPP8. Molecules
CC identified as inhibiting DPP8 catalytic activity may be useful for
CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
CC in non-insulin dependent diabetes mellitus and other disorders
CC involving glucose intolerance, enhancing mucosal regeneration and
CC as immunosuppressants.
XX
XX Sequence 360 AA;
SQ
Query Match 100.0%; Score 1933; DB 22; Length 360;
Best Local Similarity 100.0%; Pred. No. 5.3e-198;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEDARSAGVATFVQOEEDRYSGYWMCCKAETTSQGGKILILVEENDESEVEIITHVSP 60
DB 1 EEDARSAGVATFVQOEEDRYSGYWMCCKAETTSQGGKILILVEENDESEVEIITHVSP 60
QY MLETRRADSFRRYKGTANPKVTFKMSIIMIDABGRITVDVRRRLVYEGTKDSPLBHL 120
DB MLETRRADSFRRYKGTANPKVTFKMSIIMIDABGRITVDVRRRLVYEGTKDSPLBHL 120
QY 121 YVSVYVNGEYTRLTLDRGSHSCCSQCHDFFISKYSNQKNPHCVSLYKSSPEDDPCK 180
DB 121 YVSVYVNGEYTRLTLDRGSHSCCSQCHDFFISKYSNQKNPHCVSLYKSSPEDDPCK 180
QY 181 TKFEMATILDSAGPLPDYTPPEIFSESTGFTLYGMLYKPHDLOPKKPTFLFYGGP 240
DB 181 TKFEMATILDSAGPLPDYTPPEIFSESTGFTLYGMLYKPHDLOPKKPTFLFYGGP 240
QY 241 QVOLVNNRFKGVKFRILNTLASLGYVVVVVDNRGSGHRGLKPEGAFFKMGQIEIDQVE 300
DB 241 QVOLVNNRFKGVKFRILNTLASLGYVVVVVDNRGSGHRGLKPEGAFFKMGQIEIDQVE 300
QY 301 GLQVLAASYRFDITLDRVGIHGMSYGYSLMALMQRSDIFRYALAGAVTLMIFDTGYT 360
DB 301 GLQVLAASYRFDITLDRVGIHGMSYGYSLMALMQRSDIFRYALAGAVTLMIFDTGYT 360
RESULT 2
AAB47187
ID AAB47187 standard; Protein; 982 AA.
AC AAB47187;
XX
XX 29-JUN-2001 (first entry)
XX
XX Human DPP8.
XX
XX Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
XX dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
XX growth hormone deficiency; glucose level; mucosal regeneration;
XX non-insulin dependent diabetes mellitus; glucose intolerance;
XX immunosuppression.
XX
XX Homo sapiens.
OS

XX
XX Key Location/Qualifiers
FH Active-site 739
FT Active-site /note= "Forms part of Ser-Asp-His catalytic triad"
FT Active-site 817
FT Active-site /note= "Forms part of Ser-Asp-His catalytic triad"
FT Active-site 849
FT Active-site /note= "Forms part of Ser-Asp-His catalytic triad"
XX
XX MO200119866-A1.
XX
XX 22-MAR-2001. *
XX
XX 11-SEP-2000; 2000WO-AUD1085.
XX
XX 10-SEP-1999; 99AU-0002762.
XX 18-FEB-2000; 2000AU-0005709.
XX
XX (UNSY) UNIV SYDNEY.
XX
XX Abdoct CA, Gorell MD;
XX
XX WPI; 2001-281520/29.
XX N-PSDB; AAC85694.
DR
DR New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT substrates, identifying inhibitors of DPP8 catalytic activity which
PT have therapeutic uses, and for detecting activated T cells -
XX
XX Claim 1; Fig 2; 78pp; English.
XX
XX This sequence represents human dipeptidyl aminopeptidase (DPP8).
CC DPP8 has substrate specificity for H-Ala-Pro-pNA, H-Gly-Pro-pNA and
CC H-Arg-Pro-pNA. Therefore, it is a prolyl oligopeptidase and a
CC dipeptidyl peptidase, because it is capable of hydrolysing the
CC peptide bond C-terminal to Pro in each of these compounds. DPP8
CC is homologous with human DPPIV. DPP8 is useful for cleaving a
CC substrate, and for detecting an activated T cell which involves
CC measuring the level of DPP8 gene expression in a T cell. The level
CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
CC in the cell. It is also useful for identifying a molecule capable
CC of inhibiting the cleavage of the substrate by DPP8. Molecules
CC identified as inhibiting DPP8 catalytic activity may be useful for
CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
CC in non-insulin dependent diabetes mellitus and other disorders
CC involving glucose intolerance, enhancing mucosal regeneration and
CC as immunosuppressants.
XX
XX Sequence 882 AA;
SQ
Query Match 95.0%; Score 1836.5; DB 22; Length 882;
Best Local Similarity 67.5%; Pred. No. 4.7e-187;
Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;
QY 1 EEDARSAGVATFVQOEEDRYSGYWMCCKAETTSQGGKILILVEENDESEVEIITHVSP 60
DB 244 EEDARSAGVATFVQOEEDRYSGYWMCCKAETTSQGGKILILVEENDESEVEIITHVSP 303
QY 61 MLETRRADSFRRYKGTANPKVTFKMSIIMIDABGRITVDVRRRLVYEGTKDSPLBHL 98
DB 304 MLETRRADSFRRYKGTANPKVTFKMSIIMIDABGRITVDVRRRLVYEGTKDSPLBHL 363
QY 99 -----
DB 364 RAGWTPBQKAMSTILDRSGTRLOIVLISPLFIPVEDDMERQRLIESVDSVTPLIIT 423
QY 99 -----
DB 424 EETDWINIHDIHVFPQSHSEIEFIIFASECKTGFRHLKITSILKESKYKSSGGLP 483
QY 99 -----VDVRRRLVYEGTKDSPLBHLIYVSVYN 127
DB 484 APSDFKPIKEIAITSGEWEVLGRHGSNIQVDEVRRRLVYEGTKDSPLBHLIYVSVYN 543

QY 128 PGEVRLTDRGSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEFMAT 187
 Db 544 PGEVRLTDRGSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEFMAT 603
 QY 188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFTYGGPOVOLVNN 247
 Db 604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFTYGGPOVOLVNN 663
 QY 248 RFGVKYFRLNTLASLGYVVVVVINDRGSGCHGLKEFGAFKYMGGQIEIDDQVEGLQYLAS 307
 Db 664 RFGVKYFRLNTLASLGYVVVVVINDRGSGCHGLKEFGAFKYMGGQIEIDDQVEGLQYLAS 723
 QY 308 RYDFIDLDRVGIGHGMSYGGYLSLMAIMQSRDIFRYAIGAPVTLMIFYDTGYT 360
 Db 724 RYDFIDLDRVGIGHGMSYGGYLSLMAIMQSRDIFRYAIGAPVTLMIFYDTGYT 776

RESULT 3

AAE24170 standard; Protein; 882 AA.

AAE24170;

23-SEP-2002 (first entry)

Human dipeptidyl peptidase 8 (DPP8) protein.

KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
 KW autoimmunity; human immunodeficiency virus; HIV infection; cytostatic;
 KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
 KW antiviral; enzyme.

OS Homo sapiens.

PN WC200234900-A1.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-AU01388.

XX 27-OCT-2000; 2000AU-0001078.

XX (UNSY) UNIV SYDNEY.

XX Abbot CA, Gorrell MD;

XX WPI; 2002-454646/48.

XX N-PSDB; AAD38956.

PT New dipeptidyl peptidase (DPP) peptidases, useful for screening
 PT inhibitors of DPP catalytic activity, which may be employed to treat
 PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 PT rejection and HIV infection -

PS Example; Fig 1; 91pp; English.

CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
 CC polymucloides encoding such proteins. The DPP peptidases are useful for
 CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
 CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 CC rejection and HIV (human immunodeficiency virus) infection. The present
 CC sequence is human DPP8 protein.

XX Sequence 882 AA;

Query Match 95.0%; Score 1836.5; DB 23; Length 882;

Best Local Similarity 67.5%; Pred. No. 4.7e-187; Indels 173; Gaps 1;

QY 1 EEDARSAGVATFVLQEEFDRYSGYWCPCAKETTPSGKILRIILEENDESEVEIHTTSP 60
 Db 244 EEDARSAGVATFVLQEEFDRYSGYWCPCAKETTPSGKILRIILEENDESEVEIHTTSP 303

QY 61 MLETRADSFYRKYTGANPKVTFMSEIMIDAEGRIL----- 98
 Db 304 MLETRADSFYRKYTGANPKVTFMSEIMIDAEGRILIDVDKELIQPEILFEGVEYIA 363
 QY 99 ----- 98
 Db 364 RAGWTEGKAMSLIDRSQTRLQVLISPELFIPEDDVMERQRLIESVPSVTPLLIY 423
 QY 99 ----- 98
 Db 424 EETDWINIHDIHFVFPQSHHEEIEFIIPASECKTGFRHLKYITISILKSKYRRSGGLP 483
 QY 99 -----VDEVRLTYVEGTDSPLEHLLYYVSYNN 127
 Db 484 APSDEKCPDKIEIAITSGEMEWLGRHGSNIQDEVRRLVYFEGTDSPLEHLLYYVSYNN 543
 QY 128 PGEVRLTDRGSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEFMAT 187
 Db 544 PGEVRLTDRGSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEFMAT 603
 QY 188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFTYGGPOVOLVNN 247
 Db 604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFTYGGPOVOLVNN 663
 QY 248 RFGVKYFRLNTLASLGYVVVVVINDRGSGCHGLKEFGAFKYMGGQIEIDDQVEGLQYLAS 307
 Db 664 RFGVKYFRLNTLASLGYVVVVVINDRGSGCHGLKEFGAFKYMGGQIEIDDQVEGLQYLAS 723
 QY 308 RYDFIDLDRVGIGHGMSYGGYLSLMAIMQSRDIFRYAIGAPVTLMIFYDTGYT 360
 Db 724 RYDFIDLDRVGIGHGMSYGGYLSLMAIMQSRDIFRYAIGAPVTLMIFYDTGYT 776

RESULT 4

ABG61591 standard; Protein; 882 AA.

ABG61591;

12-AUG-2002 (first entry)

Human DPPIV related serine protease DPP-1.

KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinnesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.

OS Homo sapiens.

PN WC200231134-A2.

XX 18-APR-2002.

XX 12-OCT-2001; 2001WO-US1874.

XX 12-OCT-2000; 2000US-240117P.

XX (FERR) FERRING BV.

XX Qi S, Aktinsanya KO, Riviere PJ, Junien J;

XX WPI; 2002-444178/47.

XX N-PSDB; ABK83322.

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain

XX Claim 17; Fig 1; 113pp; English.

CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPP-IV). The dipeptidyl peptidase IV-related proteins (DPP-IV)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPP-IV proteins.

XX Sequence 882 AA;

Query Match 95.0%; Score 1836.5; DB 23; Length 882;
 Best Local Similarity 67.5%; Pred. No. 4.7e-187;
 Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

QY 1 EEDASAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEIIVHTSP 60
 DB 244 EEDASAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEIIVHTSP 303

QY 61 MLETRRADSPRYPKGTANPKVTFKMSIIMDAEGRII----- 98
 DB 304 MLETRRADSPRYPKGTANPKVTFKMSIIMDAEGRIIIVDKELIOPPELLFEGVEYIA 363

QY 99 ----- 98
 DB 364 RAGWTEGKYAMSLIDRSQTRLQIVLISPFLFIVEDDVMERQRLISVPSVPLIITY 423

QY 99 ----- 98
 DB 424 EETDIWINIHDFHVPQSHHEIEIFASECKTGFRHLKYKTSILKSKYKRSGGLP 483

QY 99 -----VDEVRILVYFPGTQDSPLEHLLYVSYVN 127
 DB 484 APSDEKPIKEELIATSGEMEVILGRHGSNIQVDEVRLVYFPGTQDSPLEHLLYVSYVN 543

QY 128 PGEVRLTDRGSHSCCTISOHCDPFISSKSNOKNPHCVSLYKLSPEDDPTCKTEPMAT 187
 DB 544 PGEVRLTDRGSHSCCTISOHCDPFISSKSNOKNPHCVSLYKLSPEDDPTCKTEPMAT 603

QY 188 ILDSAGPLPDYTPPEIFSFESTTGTFLYGMLYKPHDLOPKKYPTVLFIYGGPOVQLVNN 247
 DB 604 ILDSAGPLPDYTPPEIFSFESTTGTFLYGMLYKPHDLOPKKYPTVLFIYGGPOVQLVNN 663

QY 248 RFKGYKYFRRLNTLASLGIYVVVITDNRGSHRGKLFEGARKYMGQIEIDDQVEGIQYLAS 307
 DB 664 RFKGYKYFRRLNTLASLGIYVVVITDNRGSHRGKLFEGARKYMGQIEIDDQVEGIQYLAS 723

QY 308 RVDPLDLDVGIHGSYGGYLSLMLMORSDIFRYAIAAPVTLMTFYDTGYT 360
 DB 724 RVDPLDLDVGIHGSYGGYLSLMLMORSDIFRYAIAAPVTLMTFYDTGYT 776

RESULT 5
 AAU74749 standard; Protein; 882 AA.
 AAU74749;
 09-APR-2002 (first entry)
 Human protease PRPS-9 protein sequence.
 Human; protease; PRPS; gastrointestinal; Crohn's disease; cancer;
 cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;

KM inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
 KM cell proliferative disorder; developmental disorder; epilepsy;
 KM Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
 KM reproductive disorder; endometriosis.
 KM Homo sapiens.
 KM WO200196468-A2.
 KM 27-DEC-2001.
 KM 13-JUN-2001; 2001WO-US19178.
 KM 16-JUN-2000; 2000US-212336P.
 KM 22-JUN-2000; 2000US-213955P.
 KM 29-JUN-2000; 2000US-215396P.
 KM 07-JUL-2000; 2000US-216821P.
 KM 14-JUL-2000; 2000US-218946P.
 KM (INCY-) INCYTE GENOMICS INC.
 KM Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM,
 KM Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA,
 KM Walla NK, Yao MG, Lu DM, Patterson C, Tang YF, Walsh RT,
 KM Azimzal Y, Lu Y, Raskumar J, Xu Y, Reddy R, Das D, Kearney L;
 KM Kallik Da;
 KM WPI; 2002-090437/12.
 KM N-PSDB; ABK12892.
 KM Twenty one human proteases (referred to as PRPS-1 to PRPS-21), useful
 KM in the diagnosis, treatment and prevention of gastrointestinal (e.g.
 KM gastritis), cardiovascular (e.g. atherosclerosis) and cell
 KM proliferative (e.g. cancer) disorders -
 KM Claim 1; Page 140-142; 177pp; English.

QY The present invention relates to twenty one new human proteases,
 CC referred to as PRPS-1 to PRPS-21. The PRPS polynucleotides and
 CC polypeptides of the invention are useful in the diagnosis, treatment and
 CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
 CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
 CC myocardial infarction, autoimmune/inflammatory e.g. acquired
 CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
 CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
 CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
 CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
 CC endometriosis disorders. Numerous other examples of each disorder are
 CC given in the specification. The present protein sequence represents
 CC the human protease PRPS-9 protein of the invention.

XX Sequence 882 AA;

Query Match 95.0%; Score 1836.5; DB 23; Length 882;
 Best Local Similarity 67.5%; Pred. No. 4.7e-187;
 Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

QY 1 EEDASAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEIIVHTSP 60
 DB 244 EEDASAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEIIVHTSP 303

QY 61 MLETRRADSPRYPKGTANPKVTFKMSIIMDAEGRII----- 98
 DB 304 MLETRRADSPRYPKGTANPKVTFKMSIIMDAEGRIIIVDKELIOPPELLFEGVEYIA 363

QY 99 ----- 98
 DB 364 RAGWTEGKYAMSLIDRSQTRLQIVLISPFLFIVEDDVMERQRLISVPSVPLIITY 423

QY 99 ----- 98
 DB 424 EETDIWINIHDFHVPQSHHEIEIFASECKTGFRHLKYKTSILKSKYKRSGGLP 483

QY 99 -----VDEVRLTYFEGTKDSPLEHLLYVSYVN 127
 Db 484 APSDKCPKEIATITSGEWEVLGRHGSNIQYDEVARLVYFEGTKDSPLEHLLYVSYVN 543
 QY 128 PGEVTRLDRGYSHSCCISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMAT 187
 Db 544 PGEVTRLDRGYSHSCCISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMAT 603
 QY 188 IIDSAGPLPDYPPPEPFESFESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQVQLVNN 247
 Db 604 IIDSAGPLPDYPPPEPFESFESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQVQLVNN 663
 QY 248 RFGVKYFRRLNTLASLGYYVVVINDRGSGHRGLKEFGAKRYKMGQIEIDDQVEGLOYLAS 307
 Db 664 RFGVKYFRRLNTLASLGYYVVVINDRGSGHRGLKEFGAKRYKMGQIEIDDQVEGLOYLAS 723
 QY 308 RYDFIDLDRVGIHGSYGGYLSLMLMQRSDIFRVAIAGAPVTLWTFYDTGYT 360
 Db 724 RYDFIDLDRVGIHGSYGGYLSLMLMQRSDIFRVAIAGAPVTLWTFYDTGYT 776
 RESULT 6
 AAG78415 ID AAG78415 standard; Protein: 882 AA.
 AC AAG78415;
 DT 12-APR-2002 (first entry)
 DE Amino acid sequence of 21953 human prolyl oligopeptidase.
 XX
 XX
 XX
 KM 21953 prolyl oligopeptidase; antibody; proline; endopeptidase;
 KM cancer; cardiovascular disease; autoimmune disease; atopic allergy;
 KM neuronal disorder; vascular disorder; prostate disorder; cytostatic;
 KM antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
 KM diabetes mellitus; arthritis; multiple sclerosis; asthma;
 KM Grave's disease; neuronal disorder; demyelinating disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200179473-A2.
 XX
 PD 25-OCT-2001.
 XX
 PP 11-APR-2001; 2001WO-US40483.
 XX
 PR 18-APR-2000; 2000US-197508P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Meyers RA, Williamson M;
 XX
 XX WPI; 2002-034353/04.
 DR N-PsDB; AAH99934.
 PT
 PT New polypeptides 21953, member of human prolyl oligopeptidase family,
 PT useful as diagnostic targets and therapeutic agents for controlling
 PT cancer, lymphoma and leukemia -
 PS Claim 1; Page 102-103; 121pp; English.
 XX
 XX
 CC This invention relates to an isolated 21953 human prolyl
 CC oligopeptidase. Which is cytosolic, antidiabetic, antiarthritic,
 CC neuroprotective, antihypoid, dermatological, antipsoriatic,
 CC antiasthmatic, ophthalmological, antiinflammatory, nootropic,
 CC antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
 CC metabolic, cardiant, antiatherosclerotic, anorectic and
 CC metabolic in its action. Uses include gene therapy, expression or
 CC activity of 21953 protein modulator, it is useful for identifying a
 CC compound which binds to it and can be used in preventing, treating
 CC or detecting a cellular proliferative or differentiative disorder.
 CC The 21953 molecules can act as novel diagnostic targets and therapeutic
 CC agents for controlling disorders associated with the aberrant activity

CC or degradation of peptide hormones e.g., disorders associated with cell
 CC differentiation and proliferation such as cancer, immune function,
 CC reproductive, neurological and cardiovascular function. The 21953
 CC molecules are thus useful for treating and preventing cellular
 CC proliferative and differentiative disorders, haematopoietic neoplastic
 CC disorders, immune disorders such as autoimmune diseases, diabetes
 CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
 CC neuronal disorders, demyelinating diseases, vascular disorders and
 CC metabolism or pain disorders. This sequence represents the amino
 CC acid sequence of 21953 human prolyl oligopeptidase.
 XX
 SQ Sequence 882 AA;
 Query Match 95.0%; Score 1836.5; DB 23; Length 882;
 Best Local Similarity 67.5%; Pred. No. 4.7e-187;
 Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;
 QY 1 EEDARSAGVATFVLOEFPDYSYGMWCPAEETPSGKTLRLYEENDESEVLIHVTSP 60
 Db 244 EEDARSAGVATFVLOEFPDYSYGMWCPAEETPSGKTLRLYEENDESEVLIHVTSP 303
 QY 61 MLETRRADSFYRYPKTGTANPKVTFKMSIIMDAEGR1----- 98
 Db 304 MLETRRADSFYRYPKTGTANPKVTFKMSIIMDAEGR1IDVDEKELQPEILFEGVEYIA 363
 QY 99 ----- 98
 Db 364 RAGWTPCKRYAMSLIDRSQRLQIVLISPELFPVEDDWMERQRLIESVPSVTEPLIY 423
 QY 99 ----- 98
 Db 424 BETTDIWINIHDI FHVFPQSHHEE1BFI PASCKTGFRHLKYITSILKSKYKRSGG1P 483
 QY 99 -----VDEVRLTYFEGTKDSPLEHLLYVSYVN 127
 Db 484 APSDKCPKEIATITSGEWEVLGRHGSNIQYDEVARLVYFEGTKDSPLEHLLYVSYVN 543
 QY 128 PGEVTRLDRGYSHSCCISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMAT 187
 Db 544 PGEVTRLDRGYSHSCCISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMAT 603
 QY 188 IIDSAGPLPDYPPPEPFESFESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQVQLVNN 247
 Db 604 IIDSAGPLPDYPPPEPFESFESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQVQLVNN 663
 QY 248 RFGVKYFRRLNTLASLGYYVVVINDRGSGHRGLKEFGAKRYKMGQIEIDDQVEGLOYLAS 307
 Db 664 RFGVKYFRRLNTLASLGYYVVVINDRGSGHRGLKEFGAKRYKMGQIEIDDQVEGLOYLAS 723
 QY 308 RYDFIDLDRVGIHGSYGGYLSLMLMQRSDIFRVAIAGAPVTLWTFYDTGYT 360
 Db 724 RYDFIDLDRVGIHGSYGGYLSLMLMQRSDIFRVAIAGAPVTLWTFYDTGYT 776
 RESULT 7
 ABU07720 ID ABU07720 standard; Protein: 882 AA.
 XX
 XX ABU07720;
 DT 19-MAY-2003 (first entry)
 DE Human serine protease HIPHDM46.
 XX
 XX Human; enzyme; HIPHDM46; serine protease; gene therapy; osteoarthritis;
 XX serine protease activity modulation; dipeptidyl peptidase activity;
 XX musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
 KM Alzheimer's disease; paraneoplastic clear palsy; Huntington's disease;
 KM amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
 KM irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;
 KM haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;
 KM colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
 KM multiple sclerosis.

```

XX OS Homo sapiens.
XX Key Location/Qualifiers
XX Region 259..260
XX Active-site /note= "Paired glutamates of the beta propeller domain"
XX Active-site /label= Catalytic_serine_residue
XX Active-site /label= Catalytic_aspartate_residue
XX Active-site /label= Catalytic_histidine_residue
XX GB2374869-A.
XX 30-OCT-2002.
XX 22-JAN-2002; 2002GB-0001404.
XX 23-JAN-2001; 2001GB-0001760.
XX (GLAXO ) GLAXO GROUP LTD.
XX Edbrooke MR, Lewis AP;
XX MPI; 2003-150703/15.
XX N-PSDB; ABX12255.
XX Identifying modulators of serine protease activity useful for treating
XX musculoskeletal diseases, by contacting cell expressing a novel serine
XX protease polypeptide with a compound and monitoring serine protease
XX activity
XX Claim 10; Page 26-29; 38pp; English.
XX The invention relates to a method of identifying a substance that
XX modulates serine protease activity, comprising contacting a cell such as
XX a neuronal cell, lung cell, intestinal cell or a cell infected with a
XX virus, expressing a serine protease polypeptide (HHPHM 46), or its
XX variant having dipeptidyl peptidase activity, or a serine protease
XX isolated from the cell with a test substance and monitoring for serine
XX protease activity. The method is useful for identifying a substance that
XX modulates serine protease activity. A modulator of the serine protease is
XX useful in the manufacture of a medicament for treatment or prophylaxis of
XX a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
XX infection, Alzheimer's disease, paraneuronal nuclear palsy, myotonic
XX dystrophy, Huntington's disease or amyotrophic lateral sclerosis.
XX Additional disease that may be treated using modulators of the serine
XX protease include malabsorption syndromes, irritable bowel syndrome, lung
XX disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,
XX rectal polyps, small bowel tumours, colorectal tumours, anaemia,
XX dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
XX sclerosis. The present sequence represents the amino acid sequence of the
XX human serine protease HHPHM46.
XX Sequence 882 AA:
XX
XX Query Match 95.0%; Score 1836.5; DB 24; Length 882;
XX Best Local Similarity 67.5%; Pred. No. 4.7e-187; Indels 173; Gaps 1;
XX Matches 360; Conservative 0; Mismatches 0;

```

```

QY 99 -----
DB 424 EETDIDMINHDIRHVPQSHBEIEPIFASECKTGFRHLKYITSLKESKXKSSGCLP 483
QY 99 -----VDEVRRLVYFEGTKDSPLEHLLYVVSYYN 127
DB 484 APSDFKCPKIEIAITSGEMEVLRHGSNIQVDEVRRLVYFEGTKDSPLEHLLYVVSYYN 543
QY 128 PGEVTRLTDRGYSHSCISQCHDFPISKYSNQKPHCVSLYKLSPEDDPTCKTEFWAT 187
DB 544 PGEVTRLTDRGYSHSCISQCHDFPISKYSNQKPHCVSLYKLSPEDDPTCKTEFWAT 603
QY 188 ILDSAGPLPDYTPPEIFSEFSTTGFTLGMVYKPHDLCPGKKYPTVLFIYGSPQVQLVNN 247
DB 604 ILDSAGPLPDYTPPEIFSEFSTTGFTLGMVYKPHDLCPGKKYPTVLFIYGSPQVQLVNN 663
QY 248 RFKGVKPYRLNTLASLGYVVVVVDNRGSCRHGLKEGAFKYMGGIEIDDQVEGLQYLAS 307
DB 664 RFKGVKPYRLNTLASLGYVVVVVDNRGSCRHGLKEGAFKYMGGIEIDDQVEGLQYLAS 723
QY 308 RYDFIDLDVRGIGHGWSYGGYLSLWALMQRSDIFRVAIAGAPVTLFIYDTGYT 360
DB 724 RYDFIDLDVRGIGHGWSYGGYLSLWALMQRSDIFRVAIAGAPVTLFIYDTGYT 776

RESULT 8
AAB93565
ID AAB93565 standard; Protein; 632 AA.
AC AAB93565;
XX 26-JUN-2001 (first entry)
DT Human protein sequence SEQ ID NO:12964.
XX Human
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX MPI; 2001-318749/34.
XX Claim 8; SEQ ID 12964; 2537bp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification, where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end

```

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX SQ Sequence 632 AA;

Query Match 79.0%; Score 1528; DB 22; Length 632;
 Best Local Similarity 57.8%; Pred. No. 3.2e-154;
 Matches 308; Conservative 0; Mismatches 1; Indels 224; Gaps 2;

QY 1 EEDARSAGVATFVLQEPFRYSYGMWCPKAEETPSGGKILRIIYENDESEVEIHHVTSP 60
 DB 45 EEDARSAGVATFVLQEPFRYSYGMWCPKAEETPSGGKILRIIYENDESEVEIHHVTSP 104
 QY 61 METRRADSFRRYKXTGTANPKVTFKMSIIMDAEGRII----- 98
 DB 105 METRRADSFRRYKXTGTANPKVTFKMSIIMDAEGRIIIVDKELIQPELIFEGVEYIA 164
 QY 99 ----- 98
 DB 165 RAGWTEPGKYAMSILDRSQTRLQIVLISPFLIPVEDDYMERQRLIESVPSVTPLIY 224
 QY 99 ----- 98
 DB 225 EETTDIWINIHDFHVPQSHHEIEIFASECKTGRHLYKTTSLIKSKYKRSGLP 284
 QY 99 -----VDEVRILYFEGTKDSPLEHLLYVSYVN 127
 DB 285 APSDFKCPIKEIATISGEMEVILGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVSYVN 344
 QY 128 PGEVTRLTRGVSHSCCISQCHDFPISKYSNQNPKCVSLYKLSPEDDPTCKTKERWAT 187
 DB 345 PGEVTRLTRGVSHSCCISQCHDFPISKYSNQNPKCVSLYKLSPEDDPTCKTKERWAT 404
 QY 188 ILDSAGPLDPYTPPELIFSFESTTGFTLYGMLYKPHLOPKKYPVLFYIGSPQVQLVNN 247
 DB 405 ILDSAGPLDPYTPPELIFSFESTTGFTLYGMLYKPHLOPKKYPVLFYIGSPQVQLVNN 464
 QY 248 REFQVYFRILNTLASLIGYVVVVDNRGSGHRLKFGAFKYMGOIEIDDQVEGLQYLAS 307
 DB 465 REFQVYFRILNTLASLIGYVVVVDNRGSGHRLKFGAFKYM----- 507
 QY 308 RYDFIDLDVGHGMSYGYLSLMMQSRDIFRVAIAGAPVTLMTFYDTGYT 360
 DB 508 -----VAIAGAPVTLMTFYDTGYT 526

RESULT 9

ABB97362 ID ABB97362 standard; Protein; 724 AA.

AC ABB97362;

DT 27-JUN-2002 (first entry)

DE Novel human protein SEQ ID NO: 630.

XX Human; anti-neuroinflammatory; anti-inflammatory; immunomodulator;
 XX anti-infectivity; cerebroprotective; cytoskeletal; rheumatic; gene therapy;
 XX neuroprotective; anti-parkinsonian; protein therapy; EST;
 XX expressed sequence tag.

OS Homo sapiens.
 XX WO200222660-A2.
 XX 21-MAR-2002.
 XX 10-SEP-2001; 2001WO-US26015.
 XX 11-SEP-2000; 2000US-0659671.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Weinman T, Drmanac RT;
 DR WPI; 2002-292408/33.
 XX N-PSDB; ABB32548.
 PT An isolated polynucleotide for treating diseases associated with its
 XX encoded polypeptide such as cancer and multiple sclerosis -
 XX Example 2; SEQ ID NO 630; 509pp; English.

CC The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a protein of the invention.

XX SQ Sequence 724 AA;

Query Match 64.9%; Score 1254.5; DB 23; Length 724;
 Best Local Similarity 48.8%; Pred. No. 7.9e-125;
 Matches 260; Conservative 0; Mismatches 0; Indels 273; Gaps 2;

QY 1 EEDARSAGVATFVLQEPFRYSYGMWCPKAEETPSGGKILRIIYENDESEVEIHHVTSP 60
 DB 186 EEDARSAGVATFVLQEPFRYSYGMWCPKAEETPSGGKILRIIYENDESEVEIHHVTSP 245
 QY 61 METRRADSFRRYKXTGTANPKVTFKMSIIMDAEGRII----- 98
 DB 246 METRRADSFRRYKXTGTANPKVTFKMSIIMDAEGRIIIVDKELIQPELIFEGVEYIA 305
 QY 99 ----- 98
 DB 306 RAGWTEPGKYAMSILDRSQTRLQIVLISPFLIPVEDDYMERQRLIESVPSVTPLIY 365
 QY 99 ----- 98
 DB 366 EETTDIWINIHDFHVPQSHHEIEIFASECKTGRHLYKTTSLIKSKYKRSGLP 425
 QY 99 -----VDEVRILYFEGTKDSPLEHLLYVSYVN 127
 DB 426 APSDFKCPIKEIATISGEMEVILGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVSYVN 485
 QY 128 PGEVTRLTRGVSHSCCISQCHDFPISKYSNQNPKCVSLYKLSPEDDPTCKTKERWAT 187
 DB 486 PGEVTRLTRGVSHSCCISQCHDFPISKYSNQNPKCVSLYKLSPEDDPTCKTKERWAT 545
 QY 188 ILDSAGPLDPYTPPELIFSFESTTGFTLYGMLYKPHLOPKKYPVLFYIGSPQVQLVNN 247
 DB 546 ILDSAGPLDPYTPPELIFSFESTTGFTLYGMLYKPHLOPKKYPVLFYIGSPQVQLVNN 599
 QY 248 REFQVYFRILNTLASLIGYVVVVDNRGSGHRLKFGAFKYMGOIEIDDQVEGLQYLAS 307
 DB 600 ----- 599
 QY 308 RYDFIDLDVGHGMSYGYLSLMMQSRDIFRVAIAGAPVTLMTFYDTGYT 360

Db 600 -----VAIAGAPVTLMTIFYDTGYT 618

RESULT 10
ABB97361

ID ABB97361 standard; Protein; 782 AA.

AC ABB97361;

DT 27-JUN-2002 (first entry)

DE Novel human protein SEQ ID NO: 629.

XX Human; antihaemic; vulnerary; antiinflammatory; immunomodulator;

KM antifertility; cerebroprotective; cytosclastic; rheumatic; gene therapy;

KW neuroprotective; antiparkinsonian; protein therapy; EST;

XX expressed sequence tag.

OS Homo sapiens.

PN WO200222660-A2.

PD 21-MAR-2002.

PF 10-SEP-2001; 2001WO-US26015.

PR 11-SEP-2000; 2000US-0659671.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-292408/33.

DR N-PSDB; ABB97361.

XX An isolated polynucleotide for treating diseases associated with its

PT encoded polypeptide such as cancer and multiple sclerosis -

XX Example 2; SEQ ID NO 629; 509pp; English.

PS The present invention provides the protein and coding sequences of 444

CC novel human proteins. These were isolated from expressed sequences tags

CC (ESTs). They can be used to stimulate cell growth, to regulate

CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth

CC e.g. in burn treatment, to regulate the immune system e.g. to treat

CC multiple sclerosis, to regulate actinin or inhibin e.g. to treat

CC infertility, to regulate haemostasis or thrombolysis e.g. to treat

CC stroke and cancer, to screen for drugs, to treat inflammatory conditions

CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.

CC Parkinson's disease. The present sequence is a protein of the invention.

XX Sequence 782 AA;

Query Match 64.9%; Score 1254.5; DB 23; Length 782;

Best Local Similarity 48.8%; Pred. No. 8.9e-125;

Matches 260; Conservative 0; Mismatches 0; Indels 273; Gaps 2;

Db 1 EEDARSGVATFVLQEFDRYSGWMPKATPTSGGKILRIIYENDESEVEIHTTSP 60

Db 244 EEDARSGVATFVLQEFDRYSGWMPKATPTSGGKILRIIYENDESEVEIHTTSP 303

QY 61 MEETRADSPRYPTGTGAPNPKVTPKSEIMIDAEGRII----- 98

Db 304 MEETRADSPRYPTGTGAPNPKVTPKSEIMIDAEGRIIIVDKELIOPFELIPGVAYIA 363

QY 99 ----- 98

Db 364 RAGMTPEGKYAWSIILDRSOTRLQIVLISPFLIPVEDVWERQRLIESVDSVTPLIY 423

QY 99 ----- 98

Db 424 EETTDIWINIHDFVFPQSHHEIEFIPASECKTGFRHLKYITSLKESKYRRSGGLP 483

QY 99 -----VDEVRLLYFEETGKOSPLEHHLYNNVSYN 127

Db 484 APSDFKCPKIEBIATISGEMEVILGRHGSNIQVDEVRLLYFEETGKOSPLEHHLYNVSYN 543

QY 128 PGEVTRLTDRGVSHSCCISOHCDPFIISKYSNQNPKCVSLYKLSPEDDPTCKEPMAT 187

Db 544 PGEVTRLTDRGVSHSCCISOHCDPFIISKYSNQNPKCVSLYKLSPEDDPTCKEPMAT 603

QY 188 ILDSAGPLPDYTPPEIIFSESTTGFTLYGMLYXPHDLQPKKXPTVLFIVGQVQLVNN 247

Db 604 ILDSAGPLPDYTPPEIIFSESTTGFTLYGMLYXPHDLQPKKXPTVLFIVGQVQLVNN 657

QY 248 RFKGVKFFRLNTLASLGYYVVVVDNRGSCHRGLKFGCAFYKXGQLEIDQVEGLQYLAS 307

Db 658 ----- 657

QY 308 RYDFIDLDRVGHGWSYGYLSIMALMQRSDIFRVAIAGAPVTLMTIFYDTGYT 360

Db 658 -----VAIAGAPVTLMTIFYDTGYT 676

RESULT 11

ID AAB24171

AC AAB24171 standard; Protein; 830 AA.

DT 23-SEP-2002 (first entry)

XX Human dipeptidyl peptidase 4 (DPP4)-like 2 protein.

KM Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;

KW autoimmunity; human immuno deficiency virus; HIV infection; cytosclastic;

KM graft rejection; antidiabetic; antiinflammatory; immunosuppressive;

XX antiviral; enzyme; DPP-4 like 2 protein.

OS Homo sapiens.

PN WO200234900-A1.

PD 02-MAY-2002.

PF 23-OCT-2001; 2001WO-AU01388.

PR 27-OCT-2000; 2000AU-0001078.

XX (UNSY) UNITV SYDNEY.

PI Abbott CA, Gorrell MD;

PI WPI; 2002-454646/48.

DR N-PSDB; AAB24171.

XX New dipeptidyl peptidase (DPP) peptidases, useful for screening

PT inhibitors of DPP catalytic activity, which may be employed to treat

PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft

PT rejection and HIV infection -

PS Disclosure; Page 82-86; 91pp; English.

CC The present invention relates to dipeptidyl peptidase (DPP) proteins and

CC polynucleotides encoding such proteins. The DPP peptidases are useful for

CC screening inhibitors of DPP catalytic activity. The inhibitors are useful

CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft

CC rejection and HIV (human immuno deficiency virus) infection. The present

CC sequence is human DPP4-like 2 protein.

Sequence 830 AA;

Query Match	63.1%	Score 1220.5;	DB 23;	Length 830;
Best Local Similarity	44.2%;	Pred. No. 4.3e-121;		
Matches 236; Conservative	54;	Mismatches 69;	Indels 175;	Gaps 2

QY	2	EDRARSQVATFVLQEEFDRKSGYMWCPKATTPSGG-KILRIIYBENDESVELIIHTSP	60
Db	201	DDRSQGVATFVLIQEEFDRFTGYMWCPTASMGSGGLTKRIILYEBVDESEVEIIVHPS	260
QY	61	MLTPRRADSFRYKYKTGTANPKVTYFKMSEIMIDAEIRII-----	98
Db	261	ALFERKTDSFRYRRTSSKNPKIALKALAEHQIDDSQIKIVSYQEKELNVPFSLPFKXYIA	320
QY	99	-----	98
Db	321	RAGTDRGKYAMAFIDRPOWLQVLPLPALFIPSTENEBORLASARAVRNQPYVY	380
QY	99	-----	98
Db	381	EBVTNWIMVHDI FYVPQPSBGEDELCLFRANECKTSGCHLXYKTAVALKQGDMSRPS	440
QY	99	-----VDEIRLVYIPSTGKDSPLBHHLYVVSUY	126
Db	441	PGDEEFCPIKEBIALTSGEMEVLAARHSGKITWNEBETGLVYFGCTKTPLEBHHLYVVSYE	500
QY	127	NPEEVLRLIDRGYSHSCCISOCHDFEISKYSNQKNPHCVSLYKLSPEDDPTCKTEFMA	186
Db	501	AGEIIVRLTPTGFSHSCSMQNFDFVSHYSVSVPVNHVYKLSGDDDELKQRFMA	560
QY	187	TILDSAGPLPYDTTPPEIISFBSSTGFTLYGMLYKPHILOQEKYPTVLYFYGBOVOLVN	246
Db	561	SMNEAASCBDPVVPEELIFPHFTRSDVRLYGMIIYKPHALOPBKQPTVLYFYGBOVOLVN	620
QY	247	NPEFGYKYERLNTLASLGYVVVVIDNRGSCHRGKLFEGCAFYKNGGOEIDOVYEGAOYLA	306
Db	621	NSFKGIKYLRNTLASLGYAAVVVIDGRSCQGRGLRFBGALKMNGGVIEIDOVYEGLOFVA	680
QY	307	SKYDFIDDLDRVGIGHGSGTGYLSMALMQSDDIRVALIAGAPVTMLIYDTGYT	360
Db	681	EKGGEIDLSRVALIHGMSYGGGLSLMGLIHKQVPEKVALIAGAPVTVMMAUYDTGYT	734

RESULT 12

ABG61592
ID ABG61592 standard; Protein; 863 AA.

AC ABG61592;

DT 12-AUG-2002 (first entry)

Human DPPIV related serine protease DPP-2.

KW	Human; intrine processes; dipeptidyl peptidase IV-related protein; DPPF
KW	DPPF; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke
KW	heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW	dyskinesia; reproductive disorder; inflammatory disorder;
KW	metabolic disorder.

OS Homo sapiens.

PN WO200231134-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.

PR 12-OCT-2000; 2000US-240117P.

PA (FERR) FERRING BV.

XX

PI Qi S, Akinsanya KO, Riviere PJ, Junien J;

DR WPI; 2002-444178/47.

DR N-PSDB; ABK83323.

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT

PS Claim 17; Fig 1; 113pp; English.

The present invention relates to the isolation of novel human serine proteases referred to as dipeptidyl peptidase IV (DPPv)-related proteins (DPPp). The dipeptidyl peptidase IV-related proteins (DPPp) and nucleic acids encoding them are useful for treating infections such as fungal, bacterial, protozoan and viral infections, particularly infections caused by human immunodeficiency virus (HIV-1 or HIV-2), pain, diabetes, precocious puberty, infertility, obesity, anorexia, bulimia, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, angina pectoris, stroke, ulcers, asthma, allergies, cancers, migraine, vomiting, psychotic and neurological disorders (e.g. anxiety, dementia, or schizophrenia), and dyskinesias. These may also be used in discovering therapeutic agents for the treatment of reproductive, inflammatory and metabolic disorders. AbG61591-AbG61612 represent human DPPp proteins.

SQ Sequence 863 AA;

Query Match	63.1%	Score 1220.5;	DB 23;	Length 863;
Best Local Similarity	44.2%;	Pred. No. 4.6e-121;		
Matches 236; Conservative	54;	Mismatches 175;	Gaps 2	

Qy	2	EDASAGATATVLOEFEPFRIGYGMWCPAETTPSGC-KILRLIYEENDESEVRI	IHTVSP	60
Db	234	DDPSAGATATVIOEFEPFRFTGYMWCPLASWEGSGCLKRLRLIYEVEDESEVRI	IVHPSP	233
Qy	61	MLETRRADSFRRPKTGTANPKVTFKXSEIMIDAEGRII	-----	98
Db	294	ALAEKRTDYSYRPRGTSGNQPKIALTKLAIEFQRTDSOGKIVSTOEKELVQPSLSLPKVEYIA	353	
Qy	99	-----	-----	98

Qy 307 SRYPFDLDRVGHGWSYGYLSMALMORSDIERVALAGAPVTLMIFYDTGYT 360
: ||||| : ||||| : : ||||| : |||||
Db 714 EKKGFIIDSRVALHGWMSYGYFLSMGLHHPQVFVAAAPAPVTWMAAYDPGYT 767

RESULT 13

ABB98134
ID ABB98134 standard; Protein; 892 AA.

XX ABB98134;
XX 17-OCT-2002 (first entry)
XX Human PMM Incyte ID 7972712CD1.
XX
XX Human; PMM; protein modification and maintenance molecule;
XX anti-convulsant; neuroprotective; nootropic; cytoskeletal; antiapoptotic;
XX antidiabetic; dermatological; antidiabetic; antiparkinsonian;
XX antianemic; antiinflammatory; antileukemic; antidiabetic; antidiabetic;
XX hepatocellular; osteoporosis; antileukemic; antipruritic; vitiligo;
XX antibacterial; fungicide; gastroenteric; antidiabetic; laxative;
XX haemostatic; anti-HIV; thrombolytic; anticoagulant; gynaecological;
XX cardiovascular; antiarteriosclerotic; hypotensive; vasodilator;
XX antitumor; antineoplastic; immunosuppressive; antiallergic; antithyroid;
XX nephrotoxic; antitumor; chymotrypsin; antithyroid; uterine;
XX ophthalmological; antiparasitic; tranquilizer; vulnerary; keratolytic;
XX auditory; antiseborrheic; antidepressant; neuroleptic; antifertility;
XX antihelminthic; prozoocides; Crohn's disease; hypertension;
XX autoimmune; inflammatory; anaemia; cell proliferative; developmental;
XX epithelial; scabies; neurological; Alzheimer's disease; reproductive;
XX ectopic pregnancy; gene therapy; vaccine; disorder;
XX dipeptidyl peptidase 8.
XX
XX Homo sapiens.
XX
XX WO200246383-A2.
XX
XX 13-JUN-2002.
XX
XX 05-DEC-2001; 2001WO-US46964.
XX
XX 08-DEC-2000; 2000US-254399P.
XX 21-DEC-2000; 2000US-257803P.
XX 05-JAN-2001; 2001US-260110P.
XX 19-JAN-2001; 2001US-262851P.
XX 25-JAN-2001; 2001US-264623P.
XX
XX (INCYTE GENOMICS INC.
XX
XX Yue H, Azimzai Y, Kallick DA, Baughn MR, Griffin JA, Swarnakar A;
XX Lal PG, Walla NK, Hafalia AJA, Gandhi AR, Au-Young J, Elliott VS;
XX Ramkumar J, Thangavelu K, Lu Y, Warren BA, Lu DAM, Lee EA;
XX Tribouley CM, Arvizu C, Delegeane AM, Yao MG, Khan FA;
XX Sanjanwala MM;
XX
XX WPI; 2002-519664/55.
XX N-PSDB; ABQ75955.
XX
XX New isolated Protein Modification and Maintenance polypeptides, useful
XX for diagnosis, and treatment of e.g. gastrointestinal disorders -
XX
XX Claim 1 (a); Page 172-174; 2000P; English.
XX
XX The invention relates to an isolated Protein Modification and Maintenance
XX (PMM) polypeptide. Polypeptides of the invention may be used in the
XX diagnosis, treatment and prevention of disorders associated with
XX decreased expression or activity of PMM. These include gastrointestinal
XX disorders (e.g. Crohn's disease), cardiovascular disorders (e.g.
XX hypertension), autoimmune/inflammatory disorders (e.g. anaemia), cell
XX proliferative disorders, developmental disorders, epithelial disorders
XX (e.g. scabies), neurological disorders (e.g. Alzheimer's disease)
XX reproductive disorders (e.g. ectopic pregnancy), and in gene therapy or a
XX vaccine for such diseases. They may also be used in the assessment of the
XX effects of exogenous compound on the expression of nucleic acid and amino
XX acid sequences of protein modification and maintenance molecules. The
XX current sequence represents a human PMM of the invention, which has been
XX found to have homology with human dipeptidyl peptidase 8.
XX
XX Sequence 892 AA;
XX
XX Query Match 63.1%; Score 1220.5; DB 23; Length 892;

Best Local Similarity 44.2%; Pred. No. 4-8e-121;
Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;
QY 2 EDARSAGVATPVYLOEFPYSGYWCPRATTPSGG-XILRLIYEENDESEVEIHHVTS 60
DB 263 DDPKAGVATFVIOEFPFRFTGYWCPTASMGSEGLKTLRLIYEVESEVEIHHVTS 322
QY 61 MLETRARSFRPKTGCTANPKTFPMSEIMDAEGRIT----- 98
DB 323 ALLEERTDSYRPRKGSKNPKIALKLAEPQTDSDKIVSTOEKLVQPPSSLPKVEYIA 382
QY 99 ----- 98
DB 363 RAGWTRDGYAMAMFLDRPQWLQVLPLPPISTENEGRLASARAVPRNNQPYVY 442
QY 99 ----- 98
DB 443 EEVTWVWVNHDIFFPFPQSEGEDBLCFLANECKTGFCILYKTAVALKSGCYDWSSEPS 502
QY 99 -----VDVRLVYREGTKDSPLHHLYVVS 126
DB 503 PGEDEFKCPKEKIALTSGEMEVLRHSGSKLWNEETKLVFQGTDPLEHLLVVSVE 562
QY 127 NGEVTRLTDRGYSHSCCISQCHDFEISKYSNOKNPHCVSLKLSPEBDPTCKTKEFVA 186
DB 563 AAGEIVRLTTRGFHSQSCMSQNFDMFVSHYSVSFPCVHYVKLSGPDDEPLHKQPRFA 622
QY 187 TLDSAGPLPDYTPPEITSFESTGTLYMLYKPHDIQPGKTYTVLFTYGGPOVLVN 246
DB 623 SMWEASCPDPVPEIPEHFTSRSDVRLYGMITYKHALQPGKHPTVLFFVGGPQVLVN 682
QY 247 NREKGVKPRNLNTASLGVVVVVINDRSGCHRGKFEBAFKYKMOIEIDPOVEGQVYA 306
DB 683 NSFQIKYLRNLNTASLGVAVVVIDGRSCQGLKEBALKNQWQVELEDVBSGLQVFA 742
QY 307 SKYDFIDDRVGHGMSYGYSLMALQSRDIFPVALAGAVTLMIFPDYGT 360
DB 743 EKYGFIIDLSRAVHOMSYGFLSLMGLHKQVFFVALAGAVTLMAYDTGYT 796
RESULT 14 0
ABG61602
ID ABG61602 standard; Protein; 892 AA.
XX
XX ABG61602;
XX
XX 12-AUG-2002 (first entry)
XX
XX Human DPRP-2 splice variant #1.
XX
XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
XX DPRP; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
XX diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
XX heart failure; hypertension; urinary retention; osteoporosis; cancer;
XX ulcer; allergy; cancer; psychotic disorder; neurological disorder;
XX dyskinetic; reproductive disorder; inflammatory disorder;
XX metabolic disorder.
XX
XX Homo sapiens.
XX
XX WO200231134-A2.
XX
XX 18-APR-2002.
XX
XX 12-OCT-2001; 2001WO-US31874.
XX
XX 12-OCT-2000; 2000US-240117P.
XX
XX (FERR) FERRING BV.
XX
XX Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
XX WPI; 2002-444178/47.
XX
XX

DR N-PSDB; ABK83333.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 XX
 XX
 PS Disclosure; Page 76-78; 113pp; English.
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPV)-related
 CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPPR proteins.
 CC
 XX
 SO Sequence 892 AA:
 Query Match 63.1%; Score 1220.5; DB 23; Length 892;
 Best Local Similarity 44.2%; Pred. No. 4.8e-121;
 Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;
 QY 2 EDASAGATFVLOEEDRISGYWCPKAEFTTSGG-KILNILENDESEVEIHTTSP 60
 DB 263 DDPSAGATFVLOEEDRISGYWCPKAEFTTSGG-KILNILENDESEVEIHTTSP 322
 QY 61 MLETRRADSPRYPTGTANPKVTFKMSIIMDAEGRII----- 98
 DB 323 ALERKTDSTYRPTGSKNPKIALKLAFTDSQKIVSTOEKELVOPFSSLFPKVEYIA 382
 QY 99 ----- 98
 DB 383 RAGWTRDGKYYAMFLDRPQOMLOLVLPALFIPSTENEBQRLASANAVERNQPVYVY 442
 QY 99 ----- 98
 DB 443 EEVTNWVINVHDIFFPPQSGEDELCELRANECKTGCHLYKTAVALKSGQYDMSSEFS 502
 QY 99 -----VDEVRILVYFEGTKDSPLEHNYVYSYV 126
 DB 503 PGDEDFKCPKEEIALNSGEVLAHRSKTIWVBEKTVLFQKIDPLEHHLYVVSYE 562
 QY 127 NPGSVTRLTDRGYSHSCISQHCDFISKYSNQNPHCVSLYKLSPEDEPTCKTEFWA 186
 DB 563 AAGEIVRLTTPGFHSCHSCMSQNFDMFVSHYSVSTPCVHYKLSGPDDELHQPFRWA 622
 QY 187 TIIDSAGPLDPTPEPFIPESESTGTGLYKGLYKPHLOQPKKPTVLYFYGGQVQVLN 246
 DB 623 SSMEEAASCPDYVPELIFHFTTRSDVRLYKMTYKPHALQPKKPTVLYFYGGQVQVLN 682
 QY 247 NREKGVYFRLNTLASLGYYVVVIDNRGSCCHRGKLFEGAFYKMGQIEIDQVNGLOLA 306
 DB 683 NSFGIKIKYRLNTLASLGYYVVVIDNRGSCCHRGKLFEGALKNQGOVEIEQVGLDFVA 742
 QY 307 SRVDFIDLRVGIHGWGSGYGLSIMALMORSIDIRVAAIAGAPVTLMFYDTGYT 360
 DB 743 EKVGFDLSRVAIHGWSYGGFLSIMGLHKQVFKVAIAGAPVTVMAVYDGYT 796

DT 12-AUG-2002 (first entry)
 XX
 DE Human DPPR-2 splice variant #3.
 XX
 XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPPR,
 KW DPPV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junten J;
 XX
 DR MPI; 2002-444178/47.
 XX
 DR N-PSDB; ABK83335.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 XX
 XX
 PS Disclosure; Page 81-84; 113pp; English.
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPV)-related
 CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPPR proteins.
 CC
 XX
 SO Sequence 892 AA:
 Query Match 63.1%; Score 1220.5; DB 23; Length 892;
 Best Local Similarity 44.2%; Pred. No. 4.8e-121;
 Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;
 QY 2 EDASAGATFVLOEEDRISGYWCPKAEFTTSGG-KILNILENDESEVEIHTTSP 60
 DB 263 DDPSAGATFVLOEEDRISGYWCPKAEFTTSGG-KILNILENDESEVEIHTTSP 322
 QY 61 MLETRRADSPRYPTGTANPKVTFKMSIIMDAEGRII----- 98
 DB 323 ALERKTDSTYRPTGSKNPKIALKLAFTDSQKIVSTOEKELVOPFSSLFPKVEYIA 382
 QY 99 ----- 98
 DB 383 RAGWTRDGKYYAMFLDRPQOMLOLVLPALFIPSTENEBQRLASANAVERNQPVYVY 442
 QY 99 ----- 98
 DB 443 EEVTNWVINVHDIFFPPQSGEDELCELRANECKTGCHLYKTAVALKSGQYDMSSEFS 502

Thu Oct 16 09:59:38 2003

us-10-070-464-7.rag

Page 12

[illegible]

Search completed: October 15, 2003, 17:10:48
Job time : 23.9534 secs

Job time : 23.9534 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:10:56 ; Search time 8.03173 Seconds
(without alignments)
1896.467 Million cell updates/sec

Title: US-10-070-464-7

Perfect score: 1933
Sequence: 1 EEDARRSGVATFVLQEEFDR.....RVAIAGAPVTLMIFYDTGYT 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptoddata/1/iaa/5A COMB.pep:*
- 2: /cgn2_6/ptoddata/1/iaa/5B COMB.pep:*
- 3: /cgn2_6/ptoddata/1/iaa/6A COMB.pep:*
- 4: /cgn2_6/ptoddata/1/iaa/6B COMB.pep:*
- 5: /cgn2_6/ptoddata/1/iaa/PTCUTS COMB.pep:*
- 6: /cgn2_6/ptoddata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086.5	56.2	310	4	US-09-794-236-4
2	347.5	18.0	771	4	US-09-462-284-2
3	315	16.3	755	5	PCT-US93-07923-3
4	315	16.3	759	5	PCT-US93-07923-2
5	315	16.3	766	1	US-08-230-491A-3
6	315	16.3	766	1	US-08-619-280A-3
7	315	16.3	766	2	US-08-940-391-3
8	315	16.3	766	4	US-09-794-236-1
9	315	16.3	766	4	US-10-002-593-6
10	257	13.3	760	1	US-08-230-491A-2
11	257	13.3	760	1	US-08-619-280A-2
12	257	13.3	760	2	US-08-940-391-2
13	204	10.6	632	2	US-09-016-080-1
14	196	10.1	657	4	US-09-355-166-1
15	166.5	8.6	622	2	US-08-664-646A-2
16	166.5	8.6	622	2	US-09-066-285-2
17	166.5	8.6	622	3	US-09-261-006-2
18	166.5	8.6	622	3	US-08-951-088-2
19	166.5	8.6	622	4	US-09-609-566-2
20	166.5	8.6	622	4	US-09-609-570-2
21	166.5	8.6	622	4	US-09-427-372-2
22	166.5	8.6	622	4	US-09-693-554-2
23	148.5	7.7	593	5	PCT-US93-07923-11
24	147.5	7.6	721	4	US-09-390-234-20
25	131.5	6.8	344	4	US-09-724-623-79
26	116	6.0	614	4	US-09-252-991A-20060
27	110.5	5.7	836	4	US-09-491-356C-21

28	105	5.4	529	4	US-09-252-991A-24711	Sequence 24711, A
29	104.5	5.4	489	1	US-08-589-893-12	Sequence 12, Appl
30	104.5	5.4	489	2	US-09-020-991-12	Sequence 12, Appl
31	104.5	5.4	489	2	US-09-062-890-12	Sequence 12, Appl
32	104.5	5.4	570	3	US-09-068-960-13	Sequence 13, Appl
33	104	5.4	275	4	US-09-198-452A-181	Sequence 181, App
34	102.5	5.3	489	4	US-08-124-674-2	Sequence 2, Appl1
35	102.5	5.3	686	3	US-09-368-166-8	Sequence 8, Appl1
36	101.5	5.3	298	4	US-09-355-166-6	Sequence 6, Appl1
37	101	5.2	318	4	US-09-355-166-16	Sequence 16, Appl
38	98.5	5.1	444	1	US-08-483-140-28	Sequence 28, Appl
39	98.5	5.1	444	2	US-08-485-938A-32	Sequence 32, Appl
40	98.5	5.1	489	1	US-08-589-893-2	Sequence 2, Appl1
41	98.5	5.1	489	1	US-08-589-893-4	Sequence 4, Appl1
42	98.5	5.1	489	1	US-08-589-893-6	Sequence 6, Appl1
43	98.5	5.1	489	1	US-08-589-893-8	Sequence 8, Appl1
44	98.5	5.1	489	1	US-08-589-893-14	Sequence 14, Appl
45	98.5	5.1	489	1	US-08-589-893-16	Sequence 16, Appl

ALIGNMENTS

```
RESULT 1
US-09-794-236-4
; Sequence 4, Application US/09794236
; Patent No. 6337069
; GENERAL INFORMATION:
; APPLICANT: Grouzmann, Eric
; APPLICANT: Lacroix, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81985/276823
; CURRENT APPLICATION NUMBER: US/09/794,236
; CURRENT FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-236-4

Query Match      56.2%; Score 1086.5; DB 4; Length 310;
Best Local Similarity 80.6%; Pred. No. 2.2e-106;
Matches 204; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

QY      108  FEGTKSPLEHMLVYVSYVNPGEVTRLTDRGYSHSCCISQHCDFPISKYSNQKNPHCVSL 167
      |||
Db      1  FEGTKSPLEHMLVYVSYVNPGEVTRLTDRGYSHSCCISQHCDFPISKYSNQKNPHCVSL 60

QY      168  YKLSPPDDPTCKTKKEFMATILDSAGPLPYTTPREIFSPSTGTGFTLYGMLYKPHDLQPG 227
      |||
Db      61  YKLSPPDDPTCKTKKEFMATILDSAGPLPYTTPREIFSPSTGTGFTLYGMLYKPHDLQPG 120

QY      228  KKTPYVLFYGGPOVQLVNNRFKGVKXYRNLTLASLGYVVVVIDNRSCHRGKLFEGAFK 287
      |||
Db      121  KKTPYVLFYGGPO----- 134

QY      288  YKKGQLEIDQVREGLOYLARSYPFIDIDRVGIGHGWSGYGLSLMALMQRSDIFPVAIAGA 347
      |||
      0
      4
      *
      0
      135  --GQLEIDQVREGLOYLARSYPFIDIDRVGIGHGWSGYGLSLMALMQRSDIFPVAIAGA 191

Db      348  PVTLMIFYDTGYT 360
      |||
      192  PVTLMIFYDTGYT 204

RESULT 2
US-09-462-284-2
; Sequence 2, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
```

```

; APPLICANT: Nestec S.A.
; APPLICANT: Monod, Michel
; APPLICANT: Doumas, Agnes
; APPLICANT: Affolter, Michael
; APPLICANT: Van Den Broek, Peter
; TITLE OF INVENTION: CLONING OF THE
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
; TITLE OF INVENTION: ASPERGILLUS ORYZAE
; FILE REFERENCE: 8265-298
; CURRENT APPLICATION NUMBER: US/09/462,284
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 771
; TYPE: FRT
; ORGANISM: Fungus
US-09-462-284-2

```

```

Query Match      18.0%; Score 347.5; DB 4; Length 771;
Best Local Similarity 25.4%; Pred. No. 1.1e-27;
Matches 124; Conservative 57; Mismatches 150; Indels 157; Gaps 16;

```

```

QY      8 GVATFVLQEEF--DRYSGYMWCPKAEFTTSGKILRIIYENDESEVEIHHVTSPLMLET 65
DB      190 GVPMIYEEELIGDRYA-LMPSDGE-----YLAYLSEFNETGVPTTYQYVMDNQ 239
QY      66 RADSF-----RYRKTGTANPKVT-----FKKSEIMT-----          91
DB      240 IAPAYPMELKIRYKVSOTNPTVTLSLNLASKVEKAPIDAFESTDLIICEVAMLDTH 299
QY      92 -----DAEGR-----          96
DB      300 TTVAAKAKENRYDOOKVAVDPTASNKATVISDRDGTGMDNLMSKYGIPKSDKDAY 359
QY      97 -----IIVDEVRRLVYFEGTKDSPLEHNL 120
DB      360 YIDISDHGMAHLVLPVSGGEPIPLTKGDWEVTSILSIDQERQVLYVLSIQHSTERHL 419
QY      121 YVSVYVNGEYTRRLTDRCYS--HSCCTISQHCDFPISKSNOKNPHCVSLYKLSPEDDPT 178
DB      420 YSVSY-STFATVPLVDDTVAAVWSASFANSYVILTYGDPVY-DELYTTNS----- 471
QY      179 CKTEFMATILDSAGPL---PDYTPPEIFSEF--STGFTLYGMLYKPHDLQPGKKYPTV 233
DB      472 --TKPL-RTITDNKAVLEQIDYALPNTTYFELPLPSGETLNMQRPLPGSSPDKKYFIL 528
QY      234 LFIYGGPQVQLVNNRFKGVKY-PRINTLASLGYVVVVVINDRSGCHRGKFEKAFKXKQ 292
DB      529 FTFYGGPQAQCEVTRKQWALNFKAVASDSELEYVTWVDNRGTGFKGRKFRSAVTRQLGL 588
QY      293 IEIDDVQVGLQYLASRYFIDLDRVGIHGMVGYGLSMALMORSDFRVALIAGAPYTLW 352
DB      589 LEADQYLAQO-QANIPMIDADHIGLWMSFGGLISKVLEKSGAFTLVITRAPVSDW 647
QY      353 IFYDTGYT 360
DB      648 RFDYSMT 655

```

```

RESULT 3
PCT-US93-07923-3
; Sequence 3, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schlossman, Stuart F.
; APPLICANT: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston

```

```

; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 755
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US93-07923-3

```

```

Query Match      16.3%; Score 315; DB 5; Length 755;
Best Local Similarity 23.4%; Pred. No. 3e-24;
Matches 120; Conservative 59; Mismatches 146; Indels 188; Gaps 22;

```

```

QY      1 EEDRSAGVATFVLQEE-FDRYSGYMWCPKAEFTTSGKILRIIYENDESEVEIHHV 56
DB      179 KEDIYNGIDTWVVEEVEFSAYSLMWSPNGTF-----LVAQGNDEVEPLIEYSF 229
QY      59 VTSPLMLETTRADSRPYRKTGTANPKVTF-----KMSIMIDAEGRITVDEVR 104
DB      230 YSDSLQPKYKTVRYPYKAGAVNPTVAFVNVNTDLSLVNMTATSIQTAPRSMILG----- 285
QY      105 LVYFEGTKDSPLEHLLVVSIVNPEVY-----RLTRGYSHSC-----CI--SQ 147
DB      286 -----DHYLDDVTWATQERISLQWLRRIQYVSVMDICDYDESSGRWNCVLARQ 333
QY      148 HCDF-----FISKYSNQNPHCV-----SLYKLSPEDE-----DPTCKTFEWA 186
DB      334 HIEMSTGWGRF-RPEEPHTLIDGNSFYKLIISNEGYRHICYFOIDKXDCPIITKGTWE 392
QY      187 TI-----          188
DB      393 VIGIEALTSPLYVVISNEYKMPGGRMLYKQLSDYTKVTCLSCENLPERCOYYSVSFSK 452
QY      189 -----LDSAGP-LPDYT-----PPEIFSF--ESTTG 211
DB      453 EAKYQLRCSGPGPLVLYLHSSVNDKGLRVLEDNSALDKMLQNVQMPSKLDFTILLNETK 512
QY      212 FTLVGMLYKPHDLQPGKKYPTVLFYGGPQVQLVNNRFKGVYFRLN---TLASLGYVV 268
DB      513 F-WYQMLPRH-FKSKKYPLILDVYAGPCQ-----KADTVRLMWATYLAATENIIV 564
QY      269 V-IDNRSGCHRGKLFEGAFKXKQIIBDQVEGLQYLASRYFIDLDRVGIHGMVGYGY 327
DB      565 ASFDGSGSYQGDGKIMAIRRLGTPEVEDQIEAARQF-SMGVVDNRKIRIATWMSYGY 623
QY      328 LSLMALMORSDFRVALIAGAPYTLMIFYDGYT 360
DB      624 VTSNVLGSGGVFRGCIAPVAPVSRWEYDVSYT 656

```

	Query Match	16.3%	Score 315;	DB 5;	Length 759;
	Best Local Similarity	23.4%;	Pred. No. 3e-24;		
	Matches	Conservative	59;	Mismatches	146; Indels 188; Gaps 22.
QY	1 EEDARSAGVATFVLQOE-FDRYSGYWMCPKAKETTPSGGKIIRILTYEENDESEVEIIT--	56			
Db	183 KEDIIVNGIDFWMYEEEFVSAYSALMWSPNGTF-----LWAQGNDDREVLPIEVSF	233			
QY	57 VTSFMLETRRADSFRYPKCTGANRYTF-----KMSEIMDAEGRIYDEVRR	104			
Db	234 YSDSLQLQPKTVRVYPPAGAVNPLVKRFVVNTDSLSSVTNATSIQTAPASMLTG----	289			
QY	105 LVVEGETKDSPLEHLLVYSVYNQGEVF-----RLTDRGYSHSC-----CI--SQ	147			
Db	290 -----DHYLCDVMTATGERISLTQWLRIQNVSWDICDYDESSGRMNCTVARQ	337			
QY	148 HCDF-----FSKYSNQKNPHCV---SLYKLSPED-----DPCTCKEFA	186			
Db	338 HIEMSTGTGWGRF-RPSEPHEFTLDGNSFYKLIISNEBGYRHICYQIDKKDCFTFKGTWE	396			
QY	187 TI-----	188			
Db	397 VIGIEALTSPYYIYISNEYKMGPMGRNLTKIQLSDTYKTVCLSCEINPERQIYVSYSRK	456			
QY	189 ----LDSAGP-LPDYT-----PELIFS--ESTTG	211			

RESULT 5
 US-08-230-491A-3
 Sequence 3, Application US/08230491A
 Patent No. 5587299
 GENERAL INFORMATION:
 APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
 APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
 TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
 TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FELSE & LYNCH
 STREET: 805 THIRD AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10022
 COMPUTER READABLE FORM:
 MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: PC-DOS
 SOFTWARE: WORDPERFECT - ASC II
 CURRENT APPLICATION NUMBER: US/08/230,491A
 APPLICATION NUMBER: US/08/230,491A
 FILING DATE: 20-APRIL-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Hanson, No. 5587299man D.
 REGISTRATION NUMBER: 30,946
 REFERENCE/DOCKET NUMBER: LUD 330
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 688-9200
 TELEFAX: (412) 838-3884
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 766 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 IS-08-230-491A-3

Query Match	16.3%	Score 315	DB 1	Length 766
Best Local Similarity	23.4%	Pred. No. 3.1e-24		
Matches 120	Conservative 59	Mismatches 146	Indels 188	Gaps 22
QY	1	BEDARSAVATFVLQEE- FDRYSGYWC CPAETTPSGKILRLIYEENDESEVEIITH----	56	
DB	190	KEDIIYNGITDWMYVEEVEFASYSALWMSPNQTF-----LAAQFNDTEVPLEIYSF	240	
QY	57	VTSPLLETRRRDSEFRYPKGTANPKTTF-----KNSELMIDAGRIIYDEVRR	104	
DB	241	YSDSESLQYPKVRRVVPYFAGAVNPYTFEFVNTDLSISVYNNATSIOTAPASKLIG----	296	
QY	105	IYVEFGTQDSEPLEHLLIYASVYNGEVT-----RLDRGRSHSC-----CI--SQ	147	
DB	297	-----DHILCDVTWATQERISLOWMLRIQYISVMDICIDYDESSGRNCCIVARQ	344	
QY	148	HCDP-----FLISKYSNOKNPHCV-----SLYKLSPED-----DPTCKTKEFWA	186	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:13:57 ; Search time 50.8676 Seconds
(without alignments)
1140.337 Million cell updates/sec

Title: US-10-070-464-7

Perfect score: 1933
Sequence: 1 EEDASAGVATVLOEFPDR.....RVALAGAPVTLNIFYDTGYT 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1836.5	95.0	882	10	US-09-976-674-1 Sequence 1, Appli
2	1836.5	95.0	882	12	US-10-054-776-2 Sequence 2, Appli
3	1836.5	95.0	882	12	US-10-170-789-38 Sequence 3, Appli
4	1220.5	63.1	863	10	US-09-976-674-3 Sequence 3, Appli
5	1220.5	63.1	892	10	US-09-976-674-23 Sequence 23, Appli
6	1220.5	63.1	892	10	US-09-976-674-27 Sequence 27, Appli
7	1201	62.1	879	10	US-09-976-674-33 Sequence 33, Appli
8	1201	62.1	879	10	US-09-976-674-35 Sequence 35, Appli
9	1199.5	62.1	658	10	US-09-976-674-19 Sequence 19, Appli
10	1199.5	62.1	661	10	US-09-976-674-11 Sequence 11, Appli
11	1199.5	62.1	690	10	US-09-976-674-7 Sequence 7, Appli
12	1132.5	58.6	832	10	US-09-976-674-29 Sequence 29, Appli
13	1132.5	58.6	832	10	US-09-976-674-31 Sequence 31, Appli
14	1113	57.6	819	10	US-09-976-674-37 Sequence 37, Appli
15	1113	57.6	819	10	US-09-976-674-39 Sequence 39, Appli

16	1086.5	56.2	310	12	US-09-993-959-4 Sequence 4, Appli
17	944	48.8	508	15	US-10-237-271-3 Sequence 3, Appli
18	931.5	48.2	613	10	US-09-976-674-21 Sequence 21, Appli
19	505	26.1	358	10	US-09-976-674-13 Sequence 13, Appli
20	368	19.0	129	11	US-09-764-891-3564 Sequence 3564, Ap
21	352	18.2	518	10	US-09-976-674-25 Sequence 25, Appli
22	343.5	17.8	432	15	US-10-237-271-4 Sequence 4, Appli
23	315	16.3	766	9	US-09-265-606-3 Sequence 3, Appli
24	315	16.3	766	12	US-09-993-959-1 Sequence 1, Appli
25	315	16.3	766	12	US-10-423-714-6 Sequence 6, Appli
26	315	16.3	766	14	US-10-002-593-6 Sequence 6, Appli
27	315	16.3	766	15	US-10-165-603-7 Sequence 7, Appli
28	308.5	15.0	710	15	US-10-156-761-10681 Sequence 10681, A
29	296.5	15.3	767	15	US-10-165-603-4 Sequence 4, Appli
30	280.5	14.5	818	12	US-10-401-437-3 Sequence 3, Appli
31	280.5	14.5	818	12	US-10-402-312-3 Sequence 3, Appli
32	280.5	14.5	818	12	US-10-402-067-3 Sequence 3, Appli
33	280.5	14.5	818	12	US-10-401-436-3 Sequence 3, Appli
34	257	13.3	760	9	US-09-265-606-2 Sequence 2, Appli
35	257	13.3	760	12	US-10-301-822-55 Sequence 55, Appli
36	257	13.3	760	15	US-10-177-293-136 Sequence 136, App
37	191.5	10.2	803	12	US-10-205-219-58 Sequence 58, Appli
38	197.5	10.2	706	10	US-09-976-674-41 Sequence 41, Appli
39	197.5	10.2	796	10	US-09-976-674-5 Sequence 5, Appli
40	197.5	10.2	796	12	US-09-870-133-2 Sequence 2, Appli
41	197.5	10.2	796	15	US-10-160-501-5 Sequence 5, Appli
42	196	10.1	657	9	US-09-950-368-1 Sequence 1, Appli
43	196	10.1	657	12	US-10-401-437-2 Sequence 2, Appli
44	196	10.1	657	12	US-10-402-312-2 Sequence 2, Appli
45	196	10.1	657	12	US-10-402-067-2 Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-976-674-1
; Sequence 1, Application US/09976674
; Patent No. US2002015843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akitsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Julien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-1

Query Match      95.0%; Score 1836.5; DB 10; Length 882;
Best Local Similarity 67.5%; Pred. No. 5.7e-179;
Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

QY 1 EEDASAGVATVLOEFPDRYSGYWCFAETTPSGGKILRIILENDESEVIHWTSP 60
    |||
DB 244 EEDASAGVATVLOEFPDRYSGYWCFAETTPSGGKILRIILENDESEVIHWTSP 303
    |||

QY 61 MEETRADSFYRPTGTANPKVTFKMSIMDAEGRII----- 98
    |||
DB 304 MEETRADSFYRPTGTANPKVTFKMSIMDAEGRIIIVDKELIQPEILFEGVEYIA 363
    |||

QY 99 ----- 98
DB 364 RAGWTPBCKYANSLIDRSQTRLQIVLISPELFIVEDDWMERORLIESVDPSTVPLITY 423
```

QY 99 -----VDEVRLLYVEGKDSPLEHLLYVSYVN 98
Db 424 EETTDIMINIHDI FHVFPQSHHEEIEFIFASECKTGFRHL YKITSILKESKYRSSGGLP 483
QY 99 -----VDEVRLLYVEGKDSPLEHLLYVSYVN 127
Db 484 APSDFKCPKEEIAITSGEWEVLGRHGSNIQVDEVRLLYVEGKDSPLEHLLYVSYVN 543
QY 128 PGEVTRLDRGYSHSCCISOHCDFISKYSNQNPHCVSLYKLSPPEDPTCKTKERWAT 187
Db 544 PGEVTRLDRGYSHSCCISOHCDFISKYSNQNPHCVSLYKLSPPEDPTCKTKERWAT 603
QY 188 ILDSAGPLDPYTPPELFSFESTTGFTLYGMLYKPHLOPKKYPVLFIYGGPQVOLLVN 247
Db 604 ILDSAGPLDPYTPPELFSFESTTGFTLYGMLYKPHLOPKKYPVLFIYGGPQVOLLVN 663
QY 248 RFKGVKXFRNLTLASLGYYVVVVDNRGSCHRGLKFEAGAFYKXGQLEIDQVGLQYLAS 307
Db 664 RFKGVKXFRNLTLASLGYYVVVVDNRGSCHRGLKFEAGAFYKXGQLEIDQVGLQYLAS 723
QY 308 RYDFIDLDVGHGWSYGYLSLMMQSRSDIFRVAIAGAPVTLMIFYDTGYT 360
Db 724 RYDFIDLDVGHGWSYGYLSLMMQSRSDIFRVAIAGAPVTLMIFYDTGYT 776

RESULT 2
US-10-054-776-2
; Sequence 2, Application US/10054776
; Publication No., US20030165818A1
; GENERAL INFORMATION:
; APPLICANT: Mark Robert Edbrooke
; APPLICANT: Alan Peter Lewis
; TITLE OF INVENTION: NOVEL PROTEIN
; FILE REFERENCE: QG104205
; CURRENT APPLICATION NUMBER: US/10/054, 776
; CURRENT FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-054-776-2

Query Match 95.0%; Score 1836.5; DB 12; Length 882;
Best Local Similarity 67.5%; Pred. No. 5,7e-179;
Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;
QY 1 EEDARSAGVATFVLOEBFDRYSGVWMCPEKATTPSGKILRIIYENDESEVEIIHVTSP 60
Db 244 EEDARSAGVATFVLOEBFDRYSGVWMCPEKATTPSGKILRIIYENDESEVEIIHVTSP 303
QY 61 MLETRRADSFRRYKGTGTANDKVTFKMSEIMIDAGRLLI-----98
Db 304 MLETRRADSFRRYKGTGTANDKVTFKMSEIMIDAGRLLIIVDKELIQPFIILFEGVEYIA 363
QY 99 -----98
Db 364 RAGWTBEGKYAMSIILDRSQTRLQIYLISPELFI PVEDDVERORLLIESVDSVTPLIIT 423
QY 99 -----98
Db 424 EETTDIMINIHDI FHVFPQSHHEEIEFIFASECKTGFRHL YKITSILKESKYRSSGGLP 483
QY 99 -----VDEVRLLYVEGKDSPLEHLLYVSYVN 127
Db 484 APSDFKCPKEEIAITSGEWEVLGRHGSNIQVDEVRLLYVEGKDSPLEHLLYVSYVN 543
QY 128 PGEVTRLDRGYSHSCCISOHCDFISKYSNQNPHCVSLYKLSPPEDPTCKTKERWAT 187
Db 544 PGEVTRLDRGYSHSCCISOHCDFISKYSNQNPHCVSLYKLSPPEDPTCKTKERWAT 603

QY 188 ILDSAGPLDPYTPPELFSFESTTGFTLYGMLYKPHLOPKKYPVLFIYGGPQVOLLVN 247
Db 604 ILDSAGPLDPYTPPELFSFESTTGFTLYGMLYKPHLOPKKYPVLFIYGGPQVOLLVN 663
QY 248 RFKGVKXFRNLTLASLGYYVVVVDNRGSCHRGLKFEAGAFYKXGQLEIDQVGLQYLAS 307
Db 664 RFKGVKXFRNLTLASLGYYVVVVDNRGSCHRGLKFEAGAFYKXGQLEIDQVGLQYLAS 723
QY 308 RYDFIDLDVGHGWSYGYLSLMMQSRSDIFRVAIAGAPVTLMIFYDTGYT 360
Db 724 RYDFIDLDVGHGWSYGYLSLMMQSRSDIFRVAIAGAPVTLMIFYDTGYT 776

RESULT 3
US-10-170-789-38
; Sequence 38, Application US/10170789
; Publication No. US20030180930A1
; GENERAL INFORMATION:
; APPLICANT: Rachel E. Meyers
; APPLICANT: Olandt, Peter J.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Williamson, Mark
; APPLICANT: Welch, Nadine
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
; FILE REFERENCE: 10448-191001
; CURRENT APPLICATION NUMBER: US/10/170, 789
; CURRENT FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 09/797, 039
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06525
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/186, 061
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 09/882, 166
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19269
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212, 078
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/934, 406
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26052
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226, 740
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/861, 801
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: PCT/US01/16549
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/205, 508
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/801, 267
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07138
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187, 454
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/829, 671
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/US01/40483
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: US 60/197, 508
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/961, 721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235, 023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045, 367
; PRIOR FILING DATE: 2001-11-07


```
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07074
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,420
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-789-38
```

```
Query Match      95.0%; Score 1836.5; DB 12; Length 882;
Best Local Similarity 67.5%; Pred. No. 5,7e-179;
Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;
```

```
QY 1 EEDASAGVATFVLQEEFDRYSGYWCPCAKETTPSGKILRIIYEENDESEVEIHHVTSF 60
DB 244 EEDASAGVATFVLQEEFDRYSGYWCPCAKETTPSGKILRIIYEENDESEVEIHHVTSF 303
QY 61 MLETRADSFRRYKGTANPKVTFKMSIMDAEGRII----- 98
DB 304 MLETRADSFRRYKGTANPKVTFKMSIMDAEGRII----- 98
QY 99 ----- 98
DB 364 RAGWTEGKYAMSLIDRSOTRQIYLISPELFIPVEDDWMERQRLIESVPSVTPLIIT 423
QY 99 ----- 98
DB 424 EETTDIMINIHDFHVFPQSHREEIEFIPASECKTGFRHLKYITSLIKESKYRSGGLP 483
QY 99 ----- 98
DB 484 ASDDFKCP1KEBALTSGEMEWLGRHGSNIQVDEVRLVYFEGTKOSP1EHLHYVSYN 543
QY 128 PSEVTRLTDRGYSHSCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMAF 187
DB 544 PSEVTRLTDRGYSHSCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMAF 603
QY 188 IIDSAGPLPDYTPPEIFSESTTGFTLYGMLYKPHDLQPKKYPVLFYGGPQVQVLNN 247
DB 604 IIDSAGPLPDYTPPEIFSESTTGFTLYGMLYKPHDLQPKKYPVLFYGGPQVQVLNN 663
QY 248 RFKGVYFRLNTLASLGYVWVVIDNRGSCHRGLKEGAFKYMGOIEIDDOVEGLOYLAS 307
DB 664 RFKGVYFRLNTLASLGYVWVVIDNRGSCHRGLKEGAFKYMGOIEIDDOVEGLOYLAS 723
QY 308 RYDFIDLDVRVGIHGMVSGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYT 360
DB 724 RYDFIDLDVRVGIHGMVSGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYT 776

RESULT 4
US-09-976-674-3
; Sequence 3, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPEITV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 863
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-3
```

```
Query Match      63.1%; Score 1220.5; DB 10; Length 863;
Best Local Similarity 44.2%; Pred. No. 7,7e-116;
Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;
```

```
QY 2 EDASAGVATFVLQEEFDRYSGYWCPCAKETTPSGKILRIIYEENDESEVEIHHVTSF 60
DB 234 DDEKAGVATFVLQEEFDRYSGYWCPCAKETTPSGKILRIIYEENDESEVEIHHVTSF 293
QY 61 MLETRADSFRRYKGTANPKVTFKMSIMDAEGRII----- 98
DB 294 ALERKTDYRIRPRTSKNPKIALKLABRQTSQKATVSQEKELVQPFSSLPKRYEIA 353
QY 99 ----- 98
DB 354 RAGWTEGKYAMSLIDRSOTRQIYLISPELFIPSTENEDQLASABAVRNVPVYV 413
QY 99 ----- 98
DB 414 EEVYTNVWVHDFYFPQSEGEDLCFLPANECKTGCHLYKVTAVLKSQGYDWSPPFS 473
QY 99 ----- 98
DB 474 PGDEFKCP1KEBALTSGEMEWLGRHGSNIQVDEVRLVYFEGTKOSP1EHLHYVSYN 533
QY 127 NPGEVTRLTDRGYSHSCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMAF 186
DB 534 AAGEIVRLTTPGSHSCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMAF 593
QY 187 TILDSAGPLPDYTPPEIFSESTTGFTLYGMLYKPHDLQPKKYPVLFYGGPQVQVLNN 246
DB 594 SMWEASCPDPDYTPPEIFSESTTGFTLYGMLYKPHDLQPKKYPVLFYGGPQVQVLNN 653
QY 247 NPFKGVYFRLNTLASLGYVWVVIDNRGSCHRGLKEGAFKYMGOIEIDDOVEGLOYLAS 306
DB 654 NPFKGVYFRLNTLASLGYVWVVIDNRGSCHRGLKEGAFKYMGOIEIDDOVEGLOYLAS 713
QY 307 RYDFIDLDVRVGIHGMVSGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYT 360
DB 714 RYDFIDLDVRVGIHGMVSGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYT 767

RESULT 5
US-09-976-674-23
; Sequence 23, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPEITV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-23

Query Match      63.1%; Score 1220.5; DB 10; Length 892;
Best Local Similarity 44.2%; Pred. No. 8,1e-116;
```

[illegible]

Query Match	63.1%	Score 1220.5	DB 10	Length 892
Best Local Similarity	44.2%	Pred. No. 8.1e-116		
Matches 236	Conservative 54	Mismatches 65	Indels 175	Gaps 2

QY	2	EDASAGVATVLOEEDPRISGYWCPALETTPSGG	-KILRIYHENDSEVEIHTVSP	60
		DDPSAGAVFVIOEEDFRFTGYWCPASNEGSGELTILRIYEVDESEVEIHTVSP		322
Db	263	DDPSAGAVFVIOEEDFRFTGYWCPASNEGSGELTILRIYEVDESEVEIHTVSP		
		MLETRRADSFERYPKGTANPVTPTKMSIIMDAEGRII		98
QY	323	ALERRKUDSYKPEPTGSKNPIATKLAFFQDSDGKIVSTOEKEIVQFSSLPKVEVIA		382

Oy	99	-----	98
Db	383	RAGMTDQKXAMAFLDPRQOQLVILLPALFIPSTENEQRLASARAVRNVOYVVY	442
Oy	99	-----	98
Db	443	EEVTNWVINWHDIFYPFPQSEGBDELCTFLRANECKTGTCGCHLYKXTAVILKSGQYDMSEFFS	502
Oy	99	-----VDEVRRLVYFEETKDSPLEHHLVVSIV	126
Db	503	PGEDEFKCPKEEIALTNGSEWELARHSGKIMWNBETLTVYFOGTQDTPLEHHLVVSYE	562
Oy	127	NPGEVTRLTRDGRYSHSCCISQHCPIFSIKSYSNQKNPHCVSLYKLSSPEDDPCTCKTEFWA	186
Db	563	AAGGIVRLTTPGFSHSCMSQNFPMFVSHSVSSTPPCVHYKJLGGPDDDPJHKQPRFWA	622
Oy	187	TIISAGPLPYTPPELFSFESTGTFLYGMLYPHDQOPKKYPTVLFYGGPOVOLVN	246
Db	623	SMEEAASCPDYPPELFFHFTIRSDVRLYGMITYPHALQPGKGFPTVLFYGGPOVOLVN	682
Oy	247	NRFGVXXFYRLNTLASLGYVVVVIDNRGSCCHRGKFEFAGFYKMGQLEIBIDQVEGLQYLA	306
Db	683	NSFGIGKXLRNLTASLGYAAVVVIDGRGSCQGRGFRFEBALKNQMGQVEIBDQVEGLQYLA	742
Oy	307	SRYPFIDLRGIGHGWSYGYGLSLMALMQRSDIRVAVIAGAPVILMIFYDPGYT	360
Db	743	EKYGFIIDLRLVAIHGWSYGYGLSLMGLIHKQVPEKVALIAGAPVITVMAVYDTGYT	796
RESULT 7			
US-09-976-674-33			
Sequence 33, Application US/09976674			
Patent No. US20020115843A1			
GENERAL INFORMATION:			
APPLICANT: Qi, Steve			
APPLICANT: Akimsanya, Karen			
APPLICANT: Riviere, Pierre			
APPLICANT: Unien, Jean-Louis			
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV			
FILE REFERENCE: 70669			
CURRENT APPLICATION NUMBER: US/09/976,674			
CURRENT FILING DATE: 2001-10-12			
PRIOR APPLICATION NUMBER: US 60/240,117			
PRIOR FILING DATE: 2000-10-12			
NUMBER OF SEQ ID NOS: 61			
SOFTWARE: PatentIn version 3.1			
SEQ ID NO 33			
LENGTH: 879			
TYPE: FRT			
ORGANISM: Homo sapiens			
US-09-976-674-33			

	Query Match	62.1%	Score 1201;	DB 10;	Length 879;
	Best Local Similarity	44.9%;	Pred. No. 7.9e-114;		
	Matches 234;	Conservative 53;	Mismatches 72;	Indels 162;	Gaps 3
Qy	2 EDARAGAAVFATGEEEDRYSGYWCCKAETTPSGG-KILRLIYEENDESEVEIIHTVTSF	60			
Db	263 DDPRASAVAFIVLIEEEDRFTGYWMCCTLASWGEGKTLILLYEEVDSEVEVIHVSP	322			
Qy	61 MLERRADSPFYPTGTANPVTFRKSEIMIDAGRIIVDRVRLL-	105			
Db	323 ALBERKRDIYRPPTGSKNRKIALKLAFQIDSGKILVSTOKELVGFSSLPFRVEYIA	382			
Qy	106 -	VY 107			
Db	383 PACTRDGXAMAEFLDRPOOMIQVLVPALFIPTSTENEBOQLASARAVPRNVQPYVVY	442			
Qy	108 FE-	109			
Db	443 EEVTVNWINVHDIFFYPFGSEGDELCEFLRANECKTGFCHLIKYTAVLKSGYDWSEEPS	502			

QY 110 -----GTDSPLEHLLVYVSVNPGSEVRLTDRCY 139
DB 503 PGDEKCPKIEIALITSGMEVLAARHSGKTOTPLEHLLVYVSYAAEIVRLTTPGF 562
QY 140 SHSCCISOHCDFFISKYNOKNPHCVSLYKLSPEDEPTCKTEFWATILDSAGPLPDYT 139
DB 563 SHSCSWSQNDMVSHTSSVSTPCVHYVLYLSDPDDPLHKQPRFWSMWEALSCPPDYV 622
QY 200 PEIIFSESTGTGLYGMLYKPHDLPCKKYPTVLYFGSPQVQVQVNNRFKGVYFRLNT 259
DB 623 PEIIFHHTRSQVRLVGMIVKPHALPGKXHPVLYFGSPQVQVQVNNRSKGIKYLELNT 662
QY 260 LASLGYVYVVIDNRGSGHRLKFEAGFKYMGQIEIDDQVEGLQYASRYDFIDLDRVGI 319
DB 683 LASLGYAVVVIDNRGSCQRLRFEAGLKNQMGQVEIEDQVEGLQFAEKYGFIDLRSVAI 742
QY 320 HGWSYGYLSLMLMQRSDIFRYAIAAGAPVLTMYDYTGXT 360
DB 743 HGWSYGGFLSLMGLIHKQVFKVAILAGAPVLTWMAVDYTGXT 783

RESULT 8

US-09-976-674-35
; Sequence 35, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-35

Query Match 62.1%; Score 1201; DB 10; Length 879;
Best Local Similarity 44.9%; Pred. No. 7.9e-114;
Matches 224; Conservative 53; Mismatches 72; Indels 162; Gaps 3;

QY 2 EDKRSAGVATFVLQEEFDRYSGYWNCPKAETTPSGG-KILRIIYENDESSEVETIHYTSP 60
DB 263 DDKRSAGVATFVLQEEFDRFTGYWNCPTASWEGSEGKILRIIYEEVDESEVEIHVPS 322
QY 61 MLETRRADSPRYPKTGTANPKVTFKMSSEIMIDAEGRIIVDEVRRL----- 105
DB 323 ALBERKTDSRYRTSGSKNPKIALKLAEPQDSQKIVSQEKEIVOPRSSLPKXVEYIA 382
QY 106 -----VY 107
DB 383 RAGWTRDGKYAMAFIDRPOWLQVLPALFIPSTENBEQRLASARAVPRNVQPVVY 442
QY 108 FE----- 109
DB 443 EEVTVNWINVDIFYPPOSSEGEDELCTLRANECKTGFCALYKVTALVKSQGYDWSSEPPS 502
QY 110 -----GTDSPLEHLLVYVSVNPGSEVRLTDRCY 139
DB 503 PGDEKCPKIEIALITSGMEVLAARHSGKTOTPLEHLLVYVSYAAEIVRLTTPGF 562
QY 140 SHSCCISOHCDFFISKYNOKNPHCVSLYKLSPEDEPTCKTEFWATILDSAGPLPDYT 199
DB 563 SHSCSWSQNDMVSHTSSVSTPCVHYVLYLSDPDDPLHKQPRFWSMWEALSCPPDYV 622
QY 200 PEIIFSESTGTGLYGMLYKPHDLPCKKYPTVLYFGSPQVQVQVNNRFKGVYFRLNT 259

DB 623 PEIIFHHTRSQVRLVGMIVKPHALPGKXHPVLYFGSPQVQVQVNNRSKGIKYLELNT 682
QY 260 LASLGYVYVVIDNRGSGHRLKFEAGFKYMGQIEIDDQVEGLQYASRYDFIDLDRVGI 319
DB 683 LASLGYAVVVIDNRGSCQRLRFEAGLKNQMGQVEIEDQVEGLQFAEKYGFIDLRSVAI 742
QY 320 HGWSYGYLSLMLMQRSDIFRYAIAAGAPVLTMYDYTGXT 360
DB 743 HGWSYGGFLSLMGLIHKQVFKVAILAGAPVLTWMAVDYTGXT 783

RESULT 9

US-09-976-674-19
; Sequence 19, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-19

Query Match 62.1%; Score 1199.5; DB 10; Length 658;
Best Local Similarity 58.0%; Pred. No. 7.3e-114;
Matches 239; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

QY 1 EEDKRSAGVATFVLQEEFDRYSGYWNCPKAETTPSGGKILRIIYENDESSEVETIHYTSP 60
DB 244 EEDKRSAGVATFVLQEEFDRYSGYWNCPKAETTPSGGKILRIIYENDESSEVETIHYTSP 303
QY 61 MLETRRADSPRYPKTGTANPKVTFKMSSEIMIDAEGRII----- 98
DB 304 MLETRRADSPRYPKTGTANPKVTFKMSSEIMIDAEGRIIIVIDKELIQFELRPGVEYIA 363
QY 99 ----- 98
DB 364 RAGWTRDGKYAMAFIDRQOTRLQVLPALFIPYEDDVMERQRLIESVPSDVTPLIY 423
QY 99 ----- 98
DB 424 EETTDWINVDIFHVPQSHHEEIEFIPASCEKTFRRHLKYITSLIKESKYRSSGGLP 483
QY 99 -----VDEVRLLYFEGTKDSPLEHLLVYVSVN 127
DB 484 APSEDFKCPKIEIALITSGMEVLAARHSGNIQVDEYARLYFEBETKOSPLEHLLVYVSVN 543
QY 128 PGSEVRLTDRCYSHSCCISOHCDFFISKYNOKNPHCVSLYKLSPEDEPTCKTEFWAT 187
DB 544 PGSEVRLTDRCYSHSCCISOHCDFFISKYNOKNPHCVSLYKLSPEDEPTCKTEFWAT 603
QY 188 ILDSAGPLPDYTPPEIIFSESTGTGLYGMLYKPHDLPCKKYPTVLYFG 239
DB 604 ILDSAGPLPDYTPPEIIFSESTGTGLYGMLYKPHDLPCKKYPTVLYFG 655

RESULT 10

US-09-976-674-11
; Sequence 11, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:

```
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 661
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-11
```

```
Query Match      62.1%; Score 1199.5; DB 10; Length 661;
Best Local Similarity 58.0%; Pred. No. 7,4e-114;
Matches 239; Conservative 0; Mismatches 0; Indels 173; Gaps 1;
```

```
QY      1 EEDARSAGVATFVLQEEFDKSGYMWCPKAEETPSGKILRLIYEENDESEVELIHTSP 60
        |||
Db      244 EEDARSAGVATFVLQEEFDKSGYMWCPKAEETPSGKILRLIYEENDESEVELIHTSP 303
QY      61 MLETRRADSRFRYPKTGTANPKVTFKXSEIMIDAGRII-----
        |||
Db      304 MLETRRADSRFRYPKTGTANPKVTFKXSEIMIDAGRIIIVDXKELIQFELLFEGVEYIA 363
QY      99 -----
        |||
Db      364 RAGWTPGKXYAMSILLDRSQRLQIVLISPFLIPVEDDVMERQRLIESVPDSVTPLIY 423
QY      99 -----
        |||
Db      424 EETTDWINIHDIHFVFPQSHHEIEFIIFASECKTGRHLKYKTSILKSKYKRSGLP 483
QY      99 -----
        |||
Db      484 APSDFKCPIKEIAITSGEWEVLGRHGSNIQVDEVRRLVFEGRKDSPLBHLHYVSVN 543
QY      128 PGEVTRLTDKRGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPPEDDPCTKKEFWAT 187
        |||
Db      544 PGEVTRLTDKRGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPPEDDPCTKKEFWAT 603
QY      188 ILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDLOPKKYPVTVLFYIGG 239
        |||
Db      604 ILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDLOPKKYPVTVLFYIGG 655
```

```
RESULT 11
US-09-976-674-7
; Sequence 7, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-7
```

```
Query Match      62.1%; Score 1199.5; DB 10; Length 690;
Best Local Similarity 58.0%; Pred. No. 7,9e-114;
Matches 239; Conservative 0; Mismatches 0; Indels 173; Gaps 1;
```

```
QY      1 EEDARSAGVATFVLQEEFDKSGYMWCPKAEETPSGKILRLIYEENDESEVELIHTSP 60
        |||
Db      244 EEDARSAGVATFVLQEEFDKSGYMWCPKAEETPSGKILRLIYEENDESEVELIHTSP 303
QY      61 MLETRRADSRFRYPKTGTANPKVTFKXSEIMIDAGRII-----
        |||
Db      304 MLETRRADSRFRYPKTGTANPKVTFKXSEIMIDAGRIIIVDXKELIQFELLFEGVEYIA 363
QY      99 -----
        |||
Db      364 RAGWTPGKXYAMSILLDRSQRLQIVLISPFLIPVEDDVMERQRLIESVPDSVTPLIY 423
QY      99 -----
        |||
Db      424 EETTDWINIHDIHFVFPQSHHEIEFIIFASECKTGRHLKYKTSILKSKYKRSGLP 483
QY      99 -----
        |||
Db      484 APSDFKCPIKEIAITSGEWEVLGRHGSNIQVDEVRRLVFEGRKDSPLBHLHYVSVN 543
QY      128 PGEVTRLTDKRGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPPEDDPCTKKEFWAT 187
        |||
Db      544 PGEVTRLTDKRGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPPEDDPCTKKEFWAT 603
QY      188 ILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDLOPKKYPVTVLFYIGG 239
        |||
Db      604 ILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDLOPKKYPVTVLFYIGG 655
```

```
RESULT 12
US-09-976-674-29
; Sequence 29, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-29
```

```
Query Match      58.6%; Score 1132.5; DB 10; Length 832;
Best Local Similarity 42.3%; Pred. No. 7,7e-107;
Matches 222; Conservative 52; Mismatches 76; Indels 175; Gaps 2;
```

```
QY      2 EEDARSAGVATFVLQEEFDKSGYMWCPKAEETPSGKILRLIYEENDESEVELIHTSP 60
        |||
Db      263 EEDARSAGVATFVLQEEFDKSGYMWCPKAEETPSGKILRLIYEENDESEVELIHTSP 322
QY      61 MLETRRADSRFRYPKTGTANPKVTFKXSEIMIDAGRII-----
        |||
Db      323 ALERKTDYRYPRYKTNPKIALKAEFQDSGKIVSQGKELVQPFSLFPRVEXIA 382
QY      99 -----
        |||
Db      383 RAGWTRDGXYAMWFLDRPQWMLQVLLPALFIPSTENEOQLASARAVRNVPYVY 442
```

QY 99 ----- 98
Db 443 EEVTNWINVHDIFFPPQSEGEDELCEFLRANECKTGFCMLYKVTAVLKSGQYDMSPEFS 502
QY 99 ----- 126
Db 503 PGEDEFCKPIKEEIALTSGEWEVLARHSGSKIWNEEKLYFQGTQKOTPLEHNLVYVSYE 562
QY 127 NPGVETRLDRGYSHSCCISQCHDFPISKYSNQNKPCHVSLYKLSPEDEDDPTCKTEFWA 186
Db 563 AAGEIVRLTTPGFHSHSCSMQNFDMFVSHSVSTPCVHYKLSGDDDPLEHNPWFMA 622
QY 187 TILDSAGPLPDYPPPELFFSESTTGFTLYGMLYKPHDLOPKKPTVLYFTYGGPQVQVLN 246
Db 623 SMMEAACPPDYPPPELFFHHTRSADVRLYGMITYKPHALOQPKKHPVLYFYGGPQVQVLN 682
QY 247 NRPFGVKYFRLNLTASIGYVVVYIDNRGSCHRGLKFEFGAFKYMGOIEIDDOVEGLQYLA 306
Db 683 NSFKGIKYLNTLNTASIGYVVVYIDNRGSCORGLRFGALKNQMGVIEIDDOVEGLQYLA 742
QY 307 SRYPFIDLDVVGHHGMSYGYLSMALMORSDFRVAIAGAPVTL 351
Db 743 EKYGFIDLSRVAIHGMSYGYLSLIMGLIHKQVFKAOPLAVPRL 787

RESULT 13
US-09-976-674-31
; Sequence 31, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPP1V
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-31

Query Match 58.6%; Score 1132.5; DB 10; Length 832;
Best Local Similarity 42.3%; Pred. No. 7,7e-107;
Matches 222; Conservative 52; Mismatches 76; Indels 175; Gaps 2;
QY 2 EDARSAGVATFVLOEEDFRYSGYWCPCKAETTPSGG-KILRIIYEENDESEVEIHTSP 60
Db 263 DDEPSAGVATFVLOEEDFRYSGYWCPCPTASWEGSEGLKILRIIYEENDESEVEIHTSP 322
QY 61 MLETRRADSPRYPKTGTANPKVTFKSEIMIDAGRII----- 98
Db 323 ALERKTDYRYPRTGSKNPKIALKLAFOQDSQKIVSQEKELVOPFSSLPKVEYIA 382
QY 99 ----- 98
Db 383 RAGNTRDGKXAMAMFLDRPOOMQLVLLPALFIPSTENEBORLASARAVRNQPYVYV 442
QY 99 ----- 98
Db 443 EEVTNWINVHDIFFPPQSEGEDELCEFLRANECKTGFCMLYKVTAVLKSGQYDMSPEFS 502
QY 99 ----- 126
Db 503 PGEDEFCKPIKEEIALTSGEWEVLARHSGSKIWNEEKLYFQGTQKOTPLEHNLVYVSYE 562
QY 127 NPGVETRLDRGYSHSCCISQCHDFPISKYSNQNKPCHVSLYKLSPEDEDDPTCKTEFWA 186

Db 563 AAGEIVRLTTPGFHSHSCSMQNFDMFVSHSVSTPCVHYKLSGDDDPLEHNPWFMA 622
QY 187 TILDSAGPLPDYPPPELFFSESTTGFTLYGMLYKPHDLOPKKPTVLYFTYGGPQVQVLN 246
Db 623 SMMEAACPPDYPPPELFFHHTRSADVRLYGMITYKPHALOQPKKHPVLYFYGGPQVQVLN 682
QY 247 NRPFGVKYFRLNLTASIGYVVVYIDNRGSCHRGLKFEFGAFKYMGOIEIDDOVEGLQYLA 306
Db 683 NSFKGIKYLNTLNTASIGYVVVYIDNRGSCORGLRFGALKNQMGVIEIDDOVEGLQYLA 742
QY 307 SRYPFIDLDVVGHHGMSYGYLSMALMORSDFRVAIAGAPVTL 351
Db 743 EKYGFIDLSRVAIHGMSYGYLSLIMGLIHKQVFKAOPLAVPRL 787

RESULT 14
US-09-976-674-37
; Sequence 37, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPP1V
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-37

Query Match 57.6%; Score 1113; DB 10; Length 819;
Best Local Similarity 43.0%; Pred. No. 7,5e-105;
Matches 220; Conservative 51; Mismatches 79; Indels 162; Gaps 3;
QY 2 EDARSAGVATFVLOEEDFRYSGYWCPCKAETTPSGG-KILRIIYEENDESEVEIHTSP 60
Db 263 DDEPSAGVATFVLOEEDFRYSGYWCPCPTASWEGSEGLKILRIIYEENDESEVEIHTSP 322
QY 61 MLETRRADSPRYPKTGTANPKVTFKSEIMIDAGRII----- 105
Db 323 ALERKTDYRYPRTGSKNPKIALKLAFOQDSQKIVSQEKELVOPFSSLPKVEYIA 382
QY 106 ----- 107
Db 383 RAGNTRDGKXAMAMFLDRPOOMQLVLLPALFIPSTENEBORLASARAVRNQPYVYV 442
QY 108 FE----- 109
Db 443 EEVTNWINVHDIFFPPQSEGEDELCEFLRANECKTGFCMLYKVTAVLKSGQYDMSPEFS 502
QY 130 ----- 139
Db 503 PGEDEFCKPIKEEIALTSGEWEVLARHSGSKIWNEEKLYFQGTQKOTPLEHNLVYVSYE 562
QY 140 SHSCCISQCHDFPISKYSNQNKPCHVSLYKLSPEDEDDPTCKTEFWATILDSAGPLPDY 199
Db 563 SHSCSMQNFDMFVSHSVSTPCVHYKLSGDDDPLEHNPWFMAAACPDPYV 622
QY 200 PPEIFSESTTGFTLYGMLYKPHDLOPKKPTVLYFTYGGPQVQVLNNRFGVKYFRLNT 259
Db 623 PPEIFHHTRSADVRLYGMITYKPHALOQPKKHPVLYFYGGPQVQVLNNRFGVKYFRLNT 682
QY 260 LASIGYVVVYIDNRGSCHRGLKFEFGAFKYMGOIEIDDOVEGLQYLASRYFIDLDRCI 319

D6 683 LASLGIYAVVIVIDRGSCQKGLRFEGLAKKQNGQVIEDQVEGLQFVAEKIGFIDLSKVAI 742

D7 320 HGMSYGGYLSLMLMQRSDIPRVAATAGAVTL 351

D8 743 HGMSYGGFSLMGLIHKPVQFKAQPLAPRRL 774

RESULT 15

```

US-09-976-674-39
: Sequence 39, Application US/09976674
: Patent No. US20020115843A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Qi, Steve
:
: APPLICANT: Akinsanya, Karen
:
: APPLICANT: Riviere, Pierre
:
: APPLICANT: Junien, Jean-Louis
:
: TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPLIV
:
: FILE REFERENCE: 70669
:
: CURRENT APPLICATION NUMBER: US/09/976,674
:
: CURRENT FILING DATE: 2001-10-12
:
: PRIOR APPLICATION NUMBER: US 60/240,117
:
: PRIOR FILING DATE: 2000-10-12
:
: NUMBER OF SEQ ID NOS: 61
:
: SOFTWARE: PatentIn version 3.1
:
: SEQ ID NO 39
:
: LENGTH: 819
:
: TYPE: FRT
:
: ORGANISM: Homo sapiens
:
: US-09-976-674-39

```

Query Match	57.6%;	Score 113;	DB 10;	Length 819;
Best Local Similarity	43.0%;	Pred. No. 7.5e-105;		
Matches 220;	Conservative 51;	Mismatches 79;	Indels 162;	Gaps 3;

QY	2	EDRASGVATPVLOEJFEFDDSGYXWCCPKAETTPSGG-KILNILEENDESVEIHTSP	60
Dd	263	DDKSGAVATPVLOEJFEFDRFTGYWCCPNASWEGSGIKTLNILEEVDSEVEIHTSP	3222
QY	61	METRRADSFYRYPKTGTANPKVTFKMSIIMDAEGRIIVDEVRLL-----	1055
Dd	323	ALERTKTDYRYPRRTSGKNPKIALTKLAEFQDTSQCKIVSTQEKELVDPFSSLPKVEYIA	3822
QY	106	-----VY	107
Dd	383	RAGWTEGKYAMAFILDRPOOMLOLVLPPLFLPSTENEBORLASAPAVRNQYVVY	4422
QY	108	FE-----	109
Dd	443	EEVTNVMVNIHFYRPFQSGEDELCELRANECKTGFCHLYKVTAVILKSGYDMSBFS	5022
QY	110	-----GKSDPLFENHLVYVSYVNPGEVTRLLDTRGY	139
Dd	503	PGEDFKCPKIEBIALTSGEWEVLARHSGKTKDPLRLNHLVYVSYEAGIYALLTPRGF	5622
QY	140	SHSCCISOHCFPIISKYSNOKNPKHCVSLYKLSPEDDPDTCKTEKFMAITLDSAGLPDYT	199
Dd	563	SHSCMSQNPDMVSHYSSVSTRPCVHYKLSGRDDPLHNPQPFMAISMEAAACRPDYV	6222
QY	200	PEPISEFESTGTTLYGMLYKPHDLOPKKPYTYVLFYGSQOVOLVNNRFKGVKFFPLNT	259
Dd	623	PEPIFFHTRSDVRLYGMKYKPHALOPGKRPYTVLFYVGQOVOLVNNFSFGIKYKTLRNT	682
QY	260	LASLGVVVVVVINDRGSCHRGLKFEAGAFKYMQOIFRIDQVGLQYLASRYFDILDRVGI	319
Dd	683	LASLGAVVVVVINDRGSCQGRGLRFBKALKNQMQVATIEQVBSGLQVFAEKYGFIDLSVAL	742
QY	320	HGWSYGGYISLMAIMORSDIFRVVYIAGAVTL	351
Dd	743	HGWSYGGYISLIMGLIHKPQVFAQPLAVPPRL	774

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:08:55 ; Search time 9.45959 Seconds
(without alignments)
3659.853 Million cell updates/sec

Title: US-10-070-464-7

Perfect score: 1933

Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLMIFYDTGYT 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	527.5	27.3	738	2	A87516
2	491.5	25.4	931	2	T32919
3	430.5	22.3	741	2	JC5142
4	369	19.1	711	2	S66261
5	315	16.3	766	1	CDH26
6	312.5	16.0	793	2	T41703
7	309	15.1	760	1	S23752
8	291.5	15.1	792	1	A39914
9	280.5	14.5	818	1	A30107
10	250	12.9	931	2	A49737
11	245	12.7	795	2	F82658
12	240	12.4	829	2	T19514
13	226.5	11.7	683	2	E87495
14	221	11.4	803	2	A41793
15	220	11.4	709	2	B82580
16	219	11.3	779	2	T25173
17	219	11.3	799	2	T25174
18	205.5	10.6	631	2	H75007
19	204	10.6	642	2	C71337
20	203	10.5	803	2	I68600
21	203	10.5	865	2	I54331
22	199	10.3	622	2	F71174
23	197.5	10.2	743	2	T37700
24	196	10.1	657	2	E70025
25	193	10.0	632	2	E75057
26	189	9.8	536	2	F90299
27	174	9.0	759	2	I38593
28	166.5	8.6	759	2	B84381
29	165	8.5	591	2	H72474

30	162.5	8.4	569	2	S74053	probable acylamino
31	149	7.7	676	2	C97775	acylamino-acid-rel
32	147.5	7.6	667	2	A87711	prolyl oligopeptid
33	147.5	7.6	721	2	T09631	probable acylamino
34	140.5	7.3	591	2	A99444	acylaminoacyl-pept
35	140.5	7.3	606	2	T35378	probable peptide h
36	139.5	7.2	745	2	T33751	hypothetical prote
37	138.5	7.2	598	2	F84199	hypothetical prote
38	137.5	7.1	654	2	AD3183	peptidase (impor
39	134.5	7.0	732	1	JC4655	peptidase (impor
40	131.5	6.8	732	1	S07624	acylaminoacyl-pept
41	130.5	6.8	572	2	F72455	probable acylamino
42	129.5	6.7	555	2	AD1794	acylase and dieste
43	127.5	6.6	555	2	AB1419	acylaminoacyl-pept
44	126.5	6.5	629	2	T15945	hypothetical prote
45	126.5	6.5	689	1	JN0585	prolyl oligopeptid

ALIGNMENTS

RESULT 1
A87516
dipeptidyl peptidase IV [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: A87516
R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Bisen, J.; Heidelberg
N. J.; Ermolaeva, M.; White, O.; Salzhery, S.L.; Shapiro, L.; Venter, J.C.; Fraser,
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: A87516
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-738 <STO>
A/Cross-references: GB:AE005673; NID:G13423647; PIDN:AAK24125.1; GSPDB:GN00148
C/Genetics: 0
A/Gene: CC2154

Query Match	27.3%	Score 527.5;	DB 2;	Length 738;
Best Local Similarity	28.5%	Pred. No. 1e-36;		
Matches 137;	Conservative 61;	Mismatches 132;	Indels 151;	Gaps 13;
QY	2	EDARSAGVATFVLQEEFDRSYGWMCPKATFTSGKILRLVENESEVEIHHVSPM	61	
DB	189	KDALSRVVAEFTVQEEIDRFTGYMSPDES-----RIVYRVDESQVIV-----	233	
QY	62	LETRRAD-----SFRYPKTGTANP-----KVTF-----	84	
DB	234	---PRADIGGATVVGQRYPGRAPVAVVDLFFVRDLASGKVTALDVGANKDIYVARVAM	290	
QY	85	-----KMS-----IMD-----	92	
DB	291	SADGKTIVYVQRLSRDOKTLDLAFDAATGAKTILDTDPHFLEVSNDPRLTDTGTFMG	350	
QY	93	-----AEGRIT-----VDEVRLLVVEGKDSPLFHLVVS	124	
DB	351	SEKDGQNHLYRYAADGKLQITKGDMPVGLGECVBARVVAIFSASITDPIERRLYVS	410	
QY	125	YNPGEVATRLTDRGYSHSCISQHCDFISRYSNQKNPHCVSLYKLSPP-----EDDPTC	179	
DB	411	YARPKPKALTSAGWMAKVAANDNGAFAGCTYSDPKTPSQTALYSADGKRVRIENKLA	470	
QY	180	KTEEFVATLDSAGRLPDYTPREIFESFTGTLYGMLYKPHDLOGKPYPLVFLYGG	239	
DB	471	EGHPYW-----PYANLPQ-----PEFGSLKADGSLHLHILKPIGFDPKAKYPAIVSYGG	523	
QY	240	PQVQLVNNRFKGVKVFRLNTLASLGGVVVVVINDRSGCHRGLEKGAFFKYGQGOEIDDOV	299	
DB	524	PHAQRVQMKWHSPE---RTYLEAGVVIKFLDRKSGNNSAKFMRLDRKLGIVVEVDQL	580	

Db 464 EAKYYQRCSGPGLPLYTLHSSVNDKRLVLEDSALDRMLQNVMPKSKLDFTILNETK 523

QY 212 FTLYGMLYKPHDLOPGKKYPTVLFYGGPOVQVLYNNRFGKVKYFRLN---TLASLGYYV 268

Db 524 F-WYQMITLPPH-FDKSKKYPDLLDVYAGPSCQ-----KADYVFRMLWATYLASTENIIV 575

QY 269 V-IDNRSSCHRLKFEBAFKYKMQIETIDQVEGLQYLARYPIDIDRGINHMSYGY 327

Db 576 ASFGKSGYQGDGIMHAINRRILGTFFVEQIEARQF-SKMGVMDNKRIALWMSYGY 634

QY 328 LSLMALMQRSDIFRVALAGAPVTLMIIFYDTGYT 360

Db 635 VTSNVLGSSGSGVFRCGIAVAPVSRMEYYDSYIT 667

RESULT 6

T41703
dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C/Accession: T41703
R/Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1998
A/Reference number: Z22011
A/Accession: T41703
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-793 <MUR>
A/Cross-references: EMBL:AL031180; PIDN:CA20138.1; GSPDB:GN00066
A/Experimental source: strain 972h-; cosmid c2E11 -chimeric
C/Genetics:
A/Gene: SPAC2E11.08
A/Map position: 1
C/Superfamily: dipeptidyl-peptidase IV

Query Match 16.2%; Score 312.5; DB 2; Length 793;
Best Local Similarity 22.9%; Pred. No. 2,3e-18;
Matches 114; Conservative 62; Mismatches 149; Indels 173; Gaps 17;

QY 8 GVATFLQEE-FDRYSGYMWCPKAEPTPSGGKILRIYENDESEVEIIHV----- 57

Db 214 GLTWMYEEELSSPTIWMSPDSD-----KIAPLKINBEIPTYHYPLATALDP 264

QY 58 TSEPLERRRADSFRYPKGTGANPKYTFKMSRI----- 89

Db 265 SLPEFDYKDMALYKPPRGNPNPSVSLFADLNSNASSNPSLWNEPLAEPVQNVLMVN 324

QY 90 -----MDAEGRII----- 98

Db 325 TSSVLVQFTNRNSTCTARLLDTELKSIHTVKECELEBGWYEVQSAKMPFLNNSLVNEN 384

QY 99 -----VDEYRRL-----VYREGYKDSF 115

Db 385 WSDGYFDLALDDVNLHAFIIPNGSSPIYLTSGAMVTDGPIHIDGFGNYYFLATLKDS 444

QY 116 LEHLIYVSVYVNPGEVTRILDRGYS--HSCCISQHCDFISKYSNOKNPHCVSLYTKSS 172

Db 445 TERHLIYVS-LDTLEIYGITDNGEDBGYSISFSPGDFYVLYNHGDPV---WQLRS 499

QY 173 PEDDPCKKEFWATILDSAGPLDYPPEI---FSFESTTGTTLXGMLYKPHDLQPGK 228

Db 500 TKDDYCLSLSTNSRLKQO---LSSITLPSVEYGLTFNDPT-FNF-MERPRPNFVNK 553

QY 229 KYRPLVLEFYGGPOVQVLYNNRFGKVKYFRMLNTASLG-----YVVVYIDNRGSHRLKX 282

Db 554 KYRPLVLEFYGGPOVQVLYNNRFGKVKYFRMLNTASLG-----YVVVYIDNRGSHRLKX 607

QY 283 EGAFKYKQGLIETIDQVEGLQYLARYPIDIDRGINHMSYGYSLMALMQRSDIFRV 342

Db 608 RYSVSRHLGEMWESYDQAGAKFVAD-LPFDVNNHVGILWMSYGYSLMALMQRSDIFRV 665

QY 343 AIAGAPVTLMIIFYDTGYT 360

Db 666 GMAVAPVTMDRLYDSYIT 683

RESULT 7

S23752
dipeptidyl-peptidase IV (EC 3.4.14.5) alpha chain - mouse

N/Alternate names: CD26 alpha subunit; THAM alpha subunit
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C/Accession: S23752; A46465; A56030
R/Marguet, D.; Bernard, A.M.; Vivier, I.; Darmoul, D.; Naquet, P.; Pierres, M.
J. Biol. Chem. 267, 2200-2208, 1992

A/Title: CDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto-
A/Reference number: S23752; MUID:92129288; PMID:1370813
A/Accession: S23752
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-760 <MAR>

A/Cross-references: EMBL:X58364
R/Vivier, I.; Marguet, D.; Naquet, P.; Bonicel, J.; Black, D.; Li, C.X.; Bernard, A.N.
J. Immunol. 147, 447-454, 1991

A/Title: Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptidase IV)
A/Reference number: A46465; MUID:91302787; PMID:1712807
A/Accession: A46465
A/Status: preliminary
A/Molecule type: protein
A/Residues: 1-20 <VIV>

A/Experimental source: M14.T thymoma cells, Swiss nu/nu
A/Note: sequence extracted from NCBI backbone (NCBI:42236)
R/Bernard, A.M.; Maitrel, M.G.; Pierres, M.; Marguet, D.
Biochemistry 33, 15204-15214, 1994

A/Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.
A/Reference number: A56030; MUID:95092780; PMID:7999781
A/Accession: A56030
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 746-760 <BER>

A/Cross-references: GB:U12620
C/Genetics:
A/Gene: CD26
C/Superfamily: dipeptidyl-peptidase IV

C/Keywords: dipeptidylpeptidase hydrolase; glycoprotein; transmembrane protein
F:213,223,315,514,679/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:624,702,734/Active site: Ser. Asp, His #status predicted

Query Match 16.0%; Score 309; DB 1; Length 760;
Best Local Similarity 25.7%; Pred. No. 4,3e-18;
Matches 128; Conservative 57; Mismatches 154; Indels 160; Gaps 23;

QY 1 EEDARSGVATFLQEE-FDRYSGYMWCPK-----AETPSGGKILRIYENDESEVE 53

Db 184 EENVYINGITDWMYEEVFGAYISALMWSPPNNTFLAYAQFNDGTGVPLIEYGF-YSDS--- 239

QY 54 IIVTSPMLTERRADSFYRYPKGTGANPKYF----- 84

Db 240 -----LQYPTWMIPIYKAGAVNPVTFVFINIDSLSSSSSAARIQIPAPASVARGD 291

QY 85 -KNSEIMIDABGNIIVDEVRLVYFEG-----TKDSPLF-HLLYV-----VSYV 126

Db 292 HYLCDDVWATEERISIQWMLRIONYSMAICDVKINLITWNCSEQOHVEMSTGWGRF 351

QY 129 NPGVTRLPN-----RGYSHSCCI---SQHCF-----RISK 155

Db 352 RPADEPHFTSDGSSFYKISDKDGKHIHCPKPKKDDCTFTKGAMEVTSIEALTSYLYY 411

QY 156 YSNQ--KNPHCVSLYK-----LSSPEDDPCK-----TKFWATILDSAGP-L 195

Db 412 ISNQYKEMPGGRNLTKIQLDRHTNVKCLSCDLNPERQYAAVSPSKAKYQQGCGPGL 471

QY 196 PDYT-----PELFSF--ESTTGTTLXGMLYKPHDLQ 225

Db 472 PLYTLRSTDKELRVLLEDSALDRMLQDVQWPSKLDFTILNETR-F-WYQMITLPPH-FD 529

Qy	226	PGKKYPTVLFYGGPQVQVLVNNRFGKVKPELN---	TLASGYVVVV-INDRGSCHHGK	281
		: : : : : : : : : : : : : : : : : :		
		: : : : : : : : : : : : : : : : : :		
Db	530	KSRYPTPLLDVYAAPCSQ-----KADSEFLNATYLASTENITVAS	FDGRSGGYGDK	583
		: : : : : : : : : : : : : : : : : :		
Qy	282	FEGAFKMGQIEIDDOVEGLQVLASRYDFDLDRVGIGHGVSYGATSLMALMQR	SDIFR	341
		::: : : : : : : : : : : : : : : : : : :		
		::: : : : : : : : : : : : : : : : : : :		
Db	584	IMHAINRRRLGLEVEDQIEAARQV-KMGFYDSKRAVMGWSYGYVTSMVLGSG	SGSVFK	642
		: : : : : : : : : : : : : : : : : :		
		: : : : : : : : : : : : : : : : : :		
Qy	342	VAIAGAPYTLNIFYDTGYT	360	
		: : : : : : : : : : : : : : : : : :		
Db	643	CGIAVAFVSRRWEYDQSVYT	661	

RESULT 8

A39914

dipeptidyl-peptidase IV (EC 3.4.14.5), membrane-bound form precursor - rat

N/Alternate names: GP110; membrane glycoprotein 110K; OX-61

N/Contains: dipeptidyl-peptidase IV, soluble form

C/Species: Rattus norvegicus (Norway rat)

C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999

C/Accession: A39914; A33315; B33315; A60730; A42203; S38949; A31781

R/Hong, W.; Doyle, D

Proc. Natl. Acad. Sci. U.S.A. 84, 7962-7966, 1987

A/Title: cDNA cloning for a bile canalicular domain-specific membrane glycoprotein of rat

A/Reference number: A39914; MUID:88068516; PMID:3479775

A/Accession: A39914

A/Molecule type: mRNA

A/Residues: 1-792 <HON>

A/Cross-references: GB:J02997; NID:G204463; PIND:AAAA1272.1; PID:G204464

R/Ogata, S.; Mismul, Y.; Ikehara, Y.

J. Biol. Chem. 264, 3596-3601, 1989

A/Title: Primary structure of rat liver dipeptidyl peptidase IV deduced from its cDNA and

A/Reference number: A33315; MUID:89123496; PMID:2563382

A/Accession: A33315

A/Molecule type: mRNA

A/Residues: 1-37, A', 39-182, 'T', 184-331, 'T', 333-351, 'C', 353-393, 'V', 395-561, 'L', 563-623.

A/Cross-references: GB:J04591; NID:G203973; PIND:AAAA1096.1; PID:G203974

A/Note: the authors translated the codon GCG for residue 38 as Arg, ACC for residue 332

A/Accession: B33315

A/Molecule type: protein

A/Residues: 1-20;35-54;427-443;505-509;511-520;530-538;593-600;602-608;618-627 <HO2>

R/McCaughan, G.W.; Wilkison, J.E.; Creswick, P.F.; Gorrell, M.D.

Hepatology 11, 534-544, 1990

A/Title: Identification of the bile canalicular cell surface molecule GP110 as the ecto-

guene.

A/Reference number: A60730; MUID:90228896; PMID:1970322

A/Accession: A60730

A/Molecule type: protein

A/Residues: 28-47,'XX',50-53,55-58 <MCC>

R/Ogata, S.; Mismul, Y.; Tsuji, E.; Takami, N.; Oda, K.; Ikehara, Y.

Biochemistry 31, 2582-2587, 1992

A/Title: Identification of the active site residues in dipeptidyl peptidase IV by affinity

A/Reference number: A42203; MUID:92190188; PMID:1347701

A/Accession: A42203

A/Molecule type: protein

A/Residues: 'R', 625-630, 'X', 632-648 <OG2>

R/iwaki-Egawa, S.; Matanabe, Y.; Fujimoto, Y.

Riol. Chem. Hoppe-Seyler 374, 973-975, 1993

A/Title: N-terminal amino acid sequence of the 60-kDa protein of rat kidney dipeptidyl F

A/Reference number: S38949; MUID:94128239; PMID:7905271

A/Accession: S38949

A/Status: Preliminary

A/Molecule type: protein

A/Residues: 261-302 <IMA>

R/Hong, W.; Doyle, D.

J. Biol. Chem. 263, 16892-16898, 1988

A/Title: Membrane orientation of rat GP110 as studied by in vitro translation.

A/Reference number: A31781; MUID:89034185; PMID:3182821

A/Accession: A31781

A/Molecule type: mRNA

A/Residues: 1-40 <HO3>

C/Comment: This protein is localized to the bile canalculus, which is the apical domain

;Superfamily: dipeptidyl-peptidase IV

C:Kewords: dipeptidylpeptidase hydrolase; glycoprotein; homodimer; liver; serine protease
F1-1792/Product: dipeptidyl-peptidase, membrane-bound form #status experimental <MATP>
F1-1782/Domain: signal sequence #link MATS #status experimental <Sig>
F1-6/Domain: intracellular #status predicted <INT>
F17-28/Domain: transmembrane #status predicted <TMN>
F129-792/Domain: extracellular #status predicted <EXT>
F129-34/Domain: propeptide #link MATS #status experimental <PRO>
F135-792/Product: dipeptidyl-peptidase, soluble form #status experimental <MATS>
F183_90_148_217_227_319_521_686/Binding site: carbohydrate (Asn) (covalent) #status predicted <BND>
F163/Active site: Ser #status experimental
F1709_741/Active site: Asp, His #status predicted

Best Local Similarity 24.1% ; Pred. No. 1,4e-16;
Matches 119; Conservative 61; Mismatches 153; Indels 161; Gaps 22;

OY 8 GVATFVIOER-FDRESGYVMCPKAETTPSGKILRIIYEENDESEVEIIHVTSPMLETRR 66
Db 0 :
195 GINWVVVEEIEFFGYSALWM-----SPNGFIYAQAOFNDGVPLIEVSFYSDSLQYPK 248
OY 67 ADSRRYKKTGTGANPKAYVF-----KMSLEIMDAEGR 96
Db ||||| :
249 TWMIPIYPAGAVNFTVKFIVNTDSLSTTTTITPMQITAPASVTTGDHYICDVAMVSEDR 308
OY IIVDEVRLAYVF-----EGTKDS-----PLEHLIYVSVY 126
Db :
309 ISLQMRLRIQNYSMVICDYDKTNLVNNCPTTGHEHTSATGWGRFRPAEPH-----FT 363
OY 127 NPGE-VTRLTDR-GISHSCCI-----SQHCDF-----FLSKYSNQ- 159
Db :
364 SDGSSPFYKIVSDKGYNHICQFOKRKEPEDCFITKGAMEVISIEALTSDYLVIISNEY 423
OY 160 -KNHCHCSLYK-----LSPEDDPTCK-----TKEFATILTDAQP-LIPYT- 199
Db :
424 KEMGGERNLYKIQLTDHTNKCLSCIDINPERCOYVSLSKEAKYVOLGCRGPPLPYTL 483
OY 200 -----PEEIFSFESETGTFT-----LYGMKYKPHDLQPGKKYP 231
Db 484 HRSIDQKELARLENBSALDMQLODVQMPSKKLDPVLTNETRFVQMLLPH-PDKSKYP 542
OY 232 TVLFITYGQPOVLVNNRFKGVKYYFRLN--TLASLGIVVVV-IDNRGSHRGLEKGAFK 287
Db :
543 LLIDIVYAGPCSQ-----KADAAFRFMATVLAISTENIIVASFGFRSGYGQDKIMAIN 596
OY 288 YKMQGIIDVOVEL-CYLASRYDFILDVGIGHGWSYGYGLSIMALMQSDIPRAVIAG 346
Db :
597 KRLLDTLVEVDIDEARPL-KMGFVDSKOVALMWSYGGVTSVWLVSQSGLVFCGLIAV 654
OY 347 APVTLMIFVDTGYT 360
Db ||||| :
655 APVSRMEYDSVYT 668

RESULT 9
A30107
dipeptidyl aminopeptidase B (EC 3.4.14.-) - Yeast (*Saccharomyces cerevisiae*)
N.Alternate names: Yproteins YHR028c
C.Species: Saccharomyces cerevisiae
C.Date: 07-Jun-1990 #sequence_revision 30-May-1997 #text_change 23-Mar-2001
C.Accession: S46780; A30107
R.Du, Z.
submitted to the EMBL Data Library, June 1994
A.Description: The sequence of *S. cerevisiae* cosmid 8082.
A.Reference number: S46773
A.Molecule type: DNA
A.Residues: 1-818 <DUZ>
A.Cross-references: EMBL:U0399; NID:g500689; PID:AAB68879.1; PIG:g500698; GSPDB:GNOC
J.Roberts, C.J.; Pohlig, G.; Rothman, J.H., Stevens, T.H.
Article: Structure, biosynthesis, and localization of dipeptidyl aminopeptidase B, and
A.Reference number: A30107; MUID:89174971; PMID:2647766
A.Accession: A30107

RESULT 11

F82858

dipeptidyl-peptidase XP0015 [imported] - Xylella fastidiosa (strain 95c)
 C/Species: Xylella fastidiosa
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C/Accession: F82858

R/anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A/Reference number: A82515; MUID:20365717; PMID:10910347

A/Note: for a complete list of authors see reference number A59328 below

A/Accession: F82858

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-795 <SIM>

A/Cross-references: GB:AE003856; GB:AE003849; NID:99104770; PIDN:AAF82828.1; GSPDB:GN001

A/Experimental source: strain 95c

R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Britton, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H

as-Neto, E.; Docena, C.; El-Dorfy, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurume, E.E.; Laiz

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasak

A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Teshabo, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A/Reference number: A59328

A/Contents: annotation

C/Genetics: X

A/Genes: XF0015

Query Match 12.7%; Score 245; DB 2; Length 795;

Best Local Similarity 27.5%; Pred. No. 1,3e-12;

Matches 98; Conservative 49; Mismatches 140; Indels 70; Gaps 16;

21 YSGYWMCPKAEETPSGGKILRLIYENDESEVEIIHVTSPMLETRADSPRYKGTANP 80

396 FNDPGLPDRNT-----LMLSEBSGSHLYTIQ-----ETGPRKQ 431

81 KATP---KMEIMDAEGRITVDEVRRLVYFEGKSPLEHMLVYSVNPAG--EYTRL 134

432 R-FFGHMEVASAPRPSADG-----RSYVF-----VCPITPGEYEVCKLD 469

135 TDGYSHSCISQHC-DFISKYSNQNPHCVSLYKLS-SPEDPTCKTEFNATIL--- 189

470 LDGSHFTLTALQGVDFDLSPDGRQ---ILVLYSSYLPRLQAVISTNGGARVITLT 525

190 -DSAGLPDYTPPELIFSFEETGP-TLYGMLYKPHDIQPKKYPVTLFIYGGPOVOLVNN 247

526 RTSTFRKQWIOPEYVQIPKSHAGVIMGKYGPASIQPKRYPIVFWFHAGAYLQVSA 585

248 RFGGVYERLNTLASL---GVVVVINDNGSCHRGKFGKAFKYMGOEIDDQVBLQ 303

586 RYP-SYFRQMHTTLVQKGYVLDDIPASAGHGDWMTALYRNNGHPELDYLDGLD 643

304 YLASRYDFIDLDRVGIGHMSYGYLSIMALMORSDFRVAIAGAPVTLFIYDPTGYT 360

644 WLVAHKGQ-DRRRAGVYGSYGFMVMAFRAPGTFKAAALRPVODMWNHET 699

Db 644 WLVAHKGQ-DRRRAGVYGSYGFMVMAFRAPGTFKAAALRPVODMWNHET 699

RESULT 12

T19514

hypothetical protein C27C12.7 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
 C/Accession: T19514

R/Thomas, K.

submitted to the EMBL Data Library, March 1996

A/Reference number: Z19134

A/Accession: T19514

A/Status: preliminary; translated from GB/EMBL/DBS

A/Molecule type: DNA

A/Residues: 1-829 <MLT>
 A/Cross-references: EMBL:Z69883; PIDN:CAA93743.1; GSPDB:GN00028; CESP:C27C12.7

A/Experimental source: clone C27C12

C/Accession: E87495

A/Genetics: CESP:C27C12.7

A/Map position: X

A/Intons: 4/3; 51/1; 123/3; 166/3; 188/2; 279/2; 392/3; 501/3; 553/1; 583/1; 606/2;

C/Superfamily: dipeptidyl-peptidase IV

Query Match

Best Local Similarity 21.1%; Score 240; DB 2; Length 829;

Matches 107; Conservative 77; Mismatches 144; Indels 178; Gaps 20;

8 GVATFVLOEE-FRYSQYWMCPKAEETPSGGKILRLIYENDESEVEIIHVTSPMLETRR 66

249 GLANMLYEBEITLSSVWVW-----SPSGRYVSYLRFDDREVNRIPLPKTID----- 296

67 ADSP-----RYPKGTANPKVTFKMEIMDAEGRITVDEV----- 102

297 -DSYVEYFELPYKAGVQNTL---VTQYIMDSNKHIVETAPNELSAANGDYVLTNK 352

103 -----RLVYFEGTQDSPLEHMLV----- 122

353 WTPRNGSDLGERELVTWAND---QNHVYPSLCNEODCVNALSPOFSIDNRQLWVSP 409

123 --VSYNPGEVTLT-----DRG--YSH-----SCCISQCHDFPT----- 153

410 KDVKGVPFETGTGLTVLPHHNDGNINYNHVAHVELDQGTGKTKWIGENFDVILVGYG 469

154 -----SKYSNQKNPHCVSLYK-----LSPEDDPT 178

470 SKIDALTFSAYGDVGCFSTYIVREAWYSNK-----TTLQKVTQDFEDCKTLGSGADPT 525

179 -----CK-----TFEPAITLIDSA-----GPLDYTPPELIF-SPESTGTGLY 215

526 GORIVVOCERFPNTRLYLVDDVDTTKKMLBEGTKAVIPFDVNMKFGKIKLPSGIDG 585

216 GMLYKPHDLOPGKKYPTVLFIYGGPOVOLVNNRFGKYPRLNTLASLGYVVVVIDRGS 275

586 YMLTANLNDGAKIPLIDITYGSPDSKQVFOKTPPTMAIQLYSQYDIAY--ARIDRGT 643

276 CHRGLKFEAGFKYKMQQIEIDQVEGLQYLASRYDFIDLRVGIGHMSYGYT-SLMALM 334

644 GGRGMVYKEAVRYKLGDAEVDTLDMIRAPINFGFIDBRILAMGMSYGGFLSKIAIK 703

335 QRSDFRVAIAGAPVTLFIYDPTGYT 360

704 DQGLVKAISIAPIVDFKTYDSAYT 729

RESULT 13

E87495

prolyl oligopeptidase family protein [imported] - Caulobacter crescentus
 C/Species: Caulobacter crescentus
 C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C/Accession: E87495

R/Nerman, W.C.; Deboy, R.T.; Dodson, R.J.; Dukin, A.S.; Gwin, M.L.; Hatt, D.H.; K

n, U.; Ernoiaeva, M.; White, O.; Salberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of Caulobacter crescentus.

A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: E87495

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-683 <STO>

A/Cross-references: GB:AE005673; NID:913423453; PIDN:AAK3961.1; GSPDB:GN00148

C/Genetics: X

A/Genes: CC1986

Query Match 11.7%; Score 226.5; DB 2; Length 683;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 16:57:30 ; Search time 5.71145 Seconds

(without alignments)
2964.152 Million cell updates/sec

Title: US-10-070-464-7

Sequence: 1 EEDARSAGVATFVLQEEPR.....RAVAGAPVTLWIFDYDGYT 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	320	16.6	765	1 DPP4_FELCA	Q9N217 felis silve
2	319	16.5	765	1 DPP4_BOVIN	P81425 bos taurus
3	315	16.3	766	1 DPP4_HUMAN	P27487 homo sapien
4	312.5	16.2	793	1 YEAB_SCHPO	Q14073 schizosacch
5	309	16.0	760	1 DPP4_MOUSE	P28843 mus musculu
6	297.5	15.4	767	1 DPP4_RAT	P14740 rattus norv
7	280.5	14.5	818	1 DAP2_YEAST	P18962 saccharomyc
8	259	13.4	761	1 SEPR_MOUSE	P97371 mus musculu
9	257	13.3	760	1 SEPR_HUMAN	Q12884 homo sapien
10	250	12.9	931	1 ST13_YEAST	P31894 saccharomyc
11	221	11.4	863	1 DPP6_BOVIN	P42659 bos taurus
12	212	11.0	859	1 DPP6_RAT	P46101 rattus norv
13	211	10.9	804	1 DPP6_MOUSE	Q92218 mus musculu
14	203	10.5	865	1 YDPE_SCHPO	P42658 homo sapien
15	197.5	10.2	853	1 YDPE_MOUSE	Q97769 schizosacch
16	196	10.1	657	1 YDPE_HUMAN	P39839 bacillus su
17	134.5	7.0	732	1 ACPH_RAT	P13798 homo sapien
18	131.5	6.8	732	1 ACPH_HUMAN	Q99426 mus musculu
19	126.5	6.5	689	1 PPCR_AERHY	Q06903 aeromonas h
20	126.5	6.5	732	1 ACPH_PIG	P19205 sus scrofa
21	118	6.1	436	1 PAPA_CAVPO	P70683 c platelet-
22	112.5	5.8	935	1 NLG3_HUMAN	Q8NF24 homo sapien
23	110.5	5.7	836	1 NLG3_RAT	Q62888 rattus norv
24	106	5.5	705	1 PPCF_FLAME	P27195 flavobacter
25	106	5.5	710	1 PPCF_PIG	P23687 sus scrofa
26	105	5.4	722	1 Y174_RICPR	P81171 rickettsia
27	103.5	5.4	705	1 PPCB_FLAME	P27028 flavobacter
28	102.5	5.3	489	1 PNBA_BACSU	P37867 bacillus su
29	100	5.2	710	1 PPCB_BOVIN	Q99426 mus musculu
30	100	5.2	710	1 PPCF_MOUSE	Q99426 mus musculu
31	98.5	5.1	444	1 PAPA_CANPA	Q28262 c platelet-
32	98	5.1	705	1 Y450_RHISN	P55656 rhicobium s
33	96	5.0	710	1 PPCB_HUMAN	P48147 homo sapien

34	95	4.9	740	1 YL31_CABEL	P34422 caenorhabdi
35	95	4.9	848	1 NLG3_RAT	Q62889 rattus norv
36	94.5	4.9	840	1 NLG3_HUMAN	Q9N217 homo sapien
37	94.5	4.9	3354	1 CADN_HUMAN	Q9N251 homo sapien
38	94	4.9	440	1 PAPA_MOUSE	Q60963 m platelet-
39	94	4.9	848	1 NLG3_HUMAN	Q9N254 homo sapien
40	93.5	4.8	419	1 CARV_CANAL	P10977 candida alb
41	93	4.8	367	1 DUS1_MOUSE	P28563 mus musculu
42	92.5	4.8	563	1 LIP2_GEOCN	P22394 geotrichum
43	92.5	4.8	843	1 NLG3_RAT	Q62765 rattus norv
44	92	4.8	816	1 NLG4_HUMAN	Q8N064 homo sapien
45	92	4.8	825	1 NLG3_MOUSE	Q8N065 mus musculu

ALIGNMENTS

```

RESULT 1
DPP4_FELCA          STANDARD;          PRT;          765 AA.
ID   DPP4_FELCA          Q9N217;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
GN   antigen CD26).
OS   DPP4 OR CD26.
OS   Felis silvestris catus (Cat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Euteria; Carnivora; Fissipedia; Felidae; Felis.
OX   NCBI_TaxID=9685;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral blood;
RX   MEDLINE=20094000; PubMed=10630304;
RA   Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
RA   Mikami T., Takahashi E.;
RT   Molecular cloning and sequencing of a cDNA encoding the feline T-cell
RT   activation antigen CD26 homologue.";
RL   Immunogenetics 50:366-368(1999).
CC   -1- FUNCTION: Removes N-terminal dipeptides sequentially from
CC   polypeptides having unsubstituted N-termini provided that the
CC   penultimate residue is proline (By similarity).
CC   -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC   Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC   Xcc is neither Pro nor hydroxyproline.
CC   -1- SUBUNIT: Homodimer (By similarity).
CC   -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC   a soluble form (By similarity).
CC   -1- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC   by proteolytic processing (By similarity).
CC   -1- SIMILARITY: Belongs to peptidase family S9B.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
EMBL; AB023952; BAA92344.1; -.
MEROPS; S09.003; -.
InterPro; IPR002469; DPPIV N term.
InterPro; IPR001375; Peptidase_S9.
InterPro; IPR002471; Prol_endopep_ser.
InterPro; IPR000379; Ser_ester_site.
Pfam; PF00930; DPPIV N term; 1.
PROSITE; PS00326; Peptidase_S9; 1.
PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
Transmembrane; Glycoprotein; Signal-anchor.
CHAIN   1..765
         DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM

```


FT CHAIN 38 765 (MDPP).
 FT DOMAIN 1 6 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
 FT TRANSMEM 7 29 (SDPP) (BY SIMILARITY).
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 30 765 EXTRACELLULAR (POTENTIAL).
 FT ACT SITE 629 629 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 707 707 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 739 739 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 765 AA; 88213 MW; 38FCE98A22B175D9 CRC64;

Query Match 16.6%; Score 320; DB 1; Length 765;
 Best Local Similarity 25.1%; Pred. No. 1.1e-19;
 Matches 125; Conservative 59; Mismatches 156; Indels 158; Gaps 22;

QY 1 EEDARAGATFLQRE-FDRYSGYWCCKAETTPSGGKILRLYENDESEVEITH--- 56
 DB 189 EENAIYNGIADWYEEIFSAISALWSPKGT-----LAVAQFDQVPLEIYSGF 239
 QY 57 VTSPLMLETRRADSFYKPTGTANPKV-----TFK 86
 DB 240 YDESEIQYMTMRIRPKAGANPYKLVITKDNLNPNATSVETPPAAMLTGDIYL 299
 QY 87 SEIMIDAGRIIVDEVRLV-----YFEGTK---DSPLEHLVY-----VSYPNG 129
 DB 300 CVTVANBERISLQWIRKIONYSVMDIRYNNSTGKISSAQEHHEMSTGQVGRFRA 359
 QY 130 EYTRILTD-----RGYSHSCCI---SQHCP-----FISKYSN 158
 DB 360 EHFHTSDGNEFYKTIISNEDGYHICFQIDKDCFTYKAMEVIGIEALTTDYIYSN 419
 QY 159 QKN--PHCVSLYKLS-----SPEDDPTCK-----TKEFATILDSAGP-L 195
 DB 420 EYKMGPGGRNLKIKIQLNDYTKVACISCELEKPE---RCQYYSVSPKSEAKYVLRCSGPEL 476
 QY 196 PPT-----PEIFSFESTTGT-----LYGMLYKPHDLP 226
 DB 477 PLYTHRSSNDELEVLBNDSALDKMLDEVQMPSKDPIILNTEKFWYQMLPPH-FDT 535
 QY 227 GKQYPTLVFIYGGPOVQVNNRFKGVKVFRLN---TLASLGVVVV-IDNRSGCHRLKF 282
 DB 536 SKKYELLIDVAGPSCQ-----KADALFRIMWATYLASTENITVASFGRSGYQSDTI 589
 QY 283 ECAFYKMGQIETIDQVEGLQYLAIRYFIDLRVGHWSYGYTLSTALMQRSDFRIV 342
 DB 590 MEAVVRRIETFEVEQIEAARQF-SKMGFVDDKRIALMGMSYGVYVTSVGLGAGSGVKRC 648
 QY 343 AIAGAPYTLMTFYDTGYT 360
 DB 649 GIAVAVSRWEYDVSVT 666

RESULT 2
 DDP4_BOVIN
 ID_DPP4_BOVIN STANDARD; PRT; 765 AA.
 AC P81425; Q8WKG8; Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
 antigen CD26) (Adenosine deaminase complexing protein) (ADCP-I)

DE (Activation molecule 3) (ACT3) (WCI0).
 GN DDP4 OR CD26.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphocytes;
 RX MEDLINE=22067734; PubMed=12073152;
 RA Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J.,
 RA Bohach G.A.;
 RT "Molecular characterization of bovine CD26 upregulated by a
 RT staphylococcal superantigen."
 RL Immunogenetics 54:216-220 (2002).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
 RC TISSUE=Thymus;
 RX MEDLINE=22021197; PubMed=11981836;
 RA Gliddon D.R., Howard C.J.;
 RT "CD26 is expressed on a restricted subpopulation of dendritic cells in
 RT vivo."
 RL Eur. J. Immunol. 32:1472-1481 (2002).
 RN [3]
 RP SEQUENCE OF 1-24.
 RC TISSUE=T-cell;
 RX MEDLINE=21482004; PubMed=11598101;
 RA Lee S.-U., Ferens W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K.,
 RA Naessens J., Bohach G.A.;
 RT "Identity of activation molecule 3 on superantigen-stimulated bovine
 RT cells is CD26."
 RL Infect. Immun. 69:7190-7193 (2001).
 RN [4]
 RP SEQUENCE OF 537-546.
 RC TISSUE=Kidney;
 RX MEDLINE=98293306; PubMed=9629661;
 RA Ben-Shoshan I., Parola A.H.;
 RT The CP-1 subunit of adenosine deaminase complexing protein from calf
 RT kidney is identical to human, mouse, and rat dipeptidyl peptidase
 RT IV."
 RL Comp. Biochem. Physiol. 119B:289-292 (1998).
 CC -!- FUNCTION: Removes N-terminal dipeptides sequentially from
 CC polypeptides having unsubstituted N-termini provided that the
 CC penultimate residue is proline. Binds and regulates the activity
 CC of ADA.
 CC -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-
 CC xcc, from a polypeptide, preferentially when Xbb is Pro, provided
 CC xcc is neither Pro nor hydroxyproline.
 CC -!- SUBUNIT: Homodimer.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
 CC a soluble form (By similarity).
 CC -!- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and
 CC several immune system tissues.
 CC -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
 CC by proteolytic processing (By similarity).
 CC -!- SIMILARITY: Belongs to peptidase family S9B.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF461806; AAL67836.1; -
 CC EMBL; AY056834; AAL23628.1; -
 CC MEROPS; S09.003; -
 CC InterPro; IPR002471; ProL_endopep_ser.
 CC InterPro; IPR000379; Ser_estrs_site.
 CC Pfam; PF00930; DPPIV_N_term; 1.
 CC Pfam; PF00326; Peptidase_S9; 1.

DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 KM Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;.
 KM Transmembrane; Glycoprotein; Signal-anchor.
 FT CHAIN 1 765 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM (MDP).
 FT CHAIN 38 765 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM (SDP) (BY SIMILARITY).
 FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 7 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 30 765 EXTRACELLULAR (POTENTIAL).
 FT ACT_SITE 629 629 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 707 707 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 739 739 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 765 AA; 88369 MM; E32165421F43E116 CRC64;
 Query Match 16.5%; Score 319; DB 1; Length 765;
 Best Local Similarity 24.5%; Pred. No. 1.3e-19;
 Matches 125; Conservative 56; Mismatches 147; Indels 182; Gaps 22;
 QY 1 EEDARSGAVTFLVQEE-FDRYSGYVWCPEKLETPPSGKILRLIYENDSEVEIIT-- 56
 DB 189 KKQVIYNGITDWWYEEVFSAYSLWMSPNSTP-----LAVAFQNDTEVPLEIYSF 239
 QY 57 VTSPLMETERRADSPRYKGTGTANPKVTFKKSEIM-----IDAEGRILVDEYRRLVYEGT 111
 DB 240 YSDBSLQPKYTKVPIPYKAGAVNPITFEFVNVNSSLSPNINATSSQQLVPPGSVLIG---- 295
 QY 112 KDSPLHEHLVYVSVNGEVT-----RLDGRYSHSCCISQHCDF-----FLSKYSNQ- 159
 DB 296 ----DHYLCDVTVWTEERISLQWLRI--QNVS-----IMDIDCDYDSTGRWISSVGRQH 344
 QY 160 -----KNHCY-----SLYKLSPP-----DPTCK--TYEPVATI 188
 DB 345 IEISTTGWVGRFPAPAEHFTSDGNSFYKIIISNEGYKHICHFQTDKRNCTFITKGAVEVI 404
 QY 189 ----- 188
 DB 405 GIEALNSDYLYISNEYKMPGARNLYKIQLNDYTKYTCSCLEINPRRCQYVSFSQEA 464
 QY 189 ----LDSAGP-LPDYT-----PPEIFSFESTTGFTL- 214
 DB 465 KYVQLRCSGPGPLPYTLHNSNNDKELALENNSDLDQVLDQVQPSKGLDFIHLHGTFW 524
 QY 215 YGMLVYEHDPQPKKYPVYLFYGGPQVQVNNRFRKVKFRIN--TLASLGVVVV-1 270
 DB 525 YQMLPPH-FDKSKKTPLELVVAGPSQ-----FADALFRINMATVYLASTENITIVASF 577
 QY 271 DNEGSRGLKEFGAEFYKMGQLEIDDOVEGLQYASRYFIDLDRVGIGWSYGGYIATL 330
 DB 578 DKGSGYQDGKIMHAIINRIRGTREVEDQIEATROP-SKMGFVDDKRIATIGWSYGGYVTS 636
 QY 331 MALMQRSDIFRVAIAGAPVTLMIFYDTGYT 360
 DB 637 MVLGAGSGVFKCGIAVAVPSKMEYDYSVYT 666
 RESULT 3
 ID DP4_HUMAN STANDARD; PRT; 766 AA.
 AC P27487;
 DT 01-AUG-1992 (Rel. 23, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (PPI03) (Adenosine deaminase complexing protein-2) (ADAP).
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NX NCBI_Taxid=9606;
 RX SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=92329551; PubMed=1352704;
 RA Msumi Y., Hayashi Y., Arakawa F., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a serine proteinase on the cell surface.";
 RN Biochim. Biophys. Acta 1131:333-336(1992).
 [2]
 RX SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95012454; PubMed=7927537;
 RA Abbott C.A., Baker E., Sutherland G.R., McCaughan G.W.;
 RT "Genomic organization, exact localization, and tissue expression of the human CD26 (dipeptidyl peptidase IV) gene.";
 RN Immunogenetics 40:331-336(1994).
 [3]
 RX SEQUENCE FROM N.A.
 RC TISSUE=Peripheral blood;
 RX MEDLINE=92325476; PubMed=1352530;
 RA Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J., Dahlberg H.N., Schlossman S.F., Morimoto C.;
 RT "Cloning and functional expression of the T cell activation antigen CD26.";
 RN J. Immunol. 149:481-486(1992).
 [4]
 RX ERRATUM.
 RX MEDLINE=93171637; PubMed=8094732;
 RA Tanaka T.;
 RN J. Immunol. 150:2090-2090(1993).
 [5]
 RX SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Srausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Kausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Hsieh F., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh N., Datchenko U., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Huilyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton B., Ketterman M., Madan A., Rodriguez S., Sanchez A., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smalins D.E., Schermer A., Schein U.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [6]
 RX SEQUENCE OF 1-551 FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=92165847; PubMed=1347043;
 RA Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C., Trolot P., Barbat A.;
 RT "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer cell lines HT-29 and Caco-2. Cloning of the complete human coding sequence and changes of dipeptidyl peptidase IV mRNA

Db 404 VIGIEALTSYDIYIISNEYMGMPGCRNLKYLQSDTYKVTCLSCENPERCQYYSVPSK 463
 QY 189 -----LDSAGP-LPDYT-----PPIFSF--ESTTG 211
 Db 464 EAKYQQLRCGSGPGLPLTLTHSSVNDKGLAVLENDNSALDKMLQVQMSKLDPLINETK 523
 QY 212 FTLYGMLYKHHDLOPGKRYTTLFTYGGPOVQVNNRFKVKYFRLN---TLASLGYVV 268
 Db 524 F-WYQWILPPEH-FDKSKKYPPLLDVYAGPCSQ-----KADYFRINWATYLASTENIIV 575
 QY 269 V-IDNMGSGHRLKFEAGAFYKMGQQLIEDOVESLQYLAIRYFIDLDRIHGWSYGGY 327
 Db 576 ASFDGSGSGYQGDIMHAINRRLGTFEVEDQIEARQF-SKMGFVDMKRIAIGWSYGGY 634
 QY 328 LSLMALMQRSDIFRVALAGAPVTLMLFYDTGYT 360
 Db 635 VTSMWLGSGSGVFKCGIAVAPVRWEYDYSVT 667

RESULT 4
 YEAR SCHPO STANDARD; PRT; 793 AA.
 ID YEAR SCHPO
 AC 014073;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative dipeptidyl aminopeptidase C2E11.08 (EC 3.4.14.-).
 GN SPACUNK4.08 OR SPAC2E11.08.
 OS Schizosaccharomyces pombe (Fission Yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NX NCB1_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Genies S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Wolkaert G., Aert R., Robben J., Gryncompiez B.,
 RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreno S., Gloux S., Lelure V., Mottier S.,
 RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shekovski G.V., Ussery D., Barrett B.G., Nurse P.,
 RA "The genome sequence of Schizosaccharomyces pombe".
 RT Nature 415:871-880(2002).
 RL Nature 415:871-880(2002).
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
 CC vacuoles (By similarity).
 CC -I- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 Db EMBL: AL031180; CA20138.1; -
 Dr PIR: T41703; T41703.
 Dr GeneDB_Spombe; SPACUNK4.08; -
 Dr InterPro; IPR002469; DPIP_V_N_term.
 Dr InterPro; IPR001375; Peptidase S9.
 Dr InterPro; IPR002471; ProL endopep. ser.
 Dr InterPro; IPR000379; Ser_estrg_site.
 Dr Pfam; PF00930; DPIP_V_N_term; 1.
 Dr Pfam; PF00326; Peptidase S9; 1.
 Dr HypoSite; PS00708; PRO_ENDOPEP_SER; 1.
 KM HypoSite; PS00708; PRO_ENDOPEP_SER; 1.
 KM HypoSite; PS00708; PRO_ENDOPEP_SER; 1.
 KM Serine protease: Transmembrane; Glycoprotein; Signal-anchor.
 FT DOMAIN 1 24
 FT TRANSMEM 25 45
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT LUMENAL (POTENTIAL).
 FT ACT SITE 647 647
 FT ACT SITE 722 722
 FT ACT SITE 755 755
 FT CARBOHYD 101 101
 FT CARBOHYD 136 136
 FT CARBOHYD 146 146
 FT CARBOHYD 246 246
 FT CARBOHYD 299 299
 FT CARBOHYD 303 303
 FT CARBOHYD 324 324
 FT CARBOHYD 336 336
 FT CARBOHYD 377 377
 FT CARBOHYD 384 384
 FT CARBOHYD 407 407
 FT CARBOHYD 535 535
 FT CARBOHYD 761 761
 SQ SEQUENCE 793 AA; 91304 MW; 20870797F231463 CRC64;
 Query Match 16.2%; Score 312.5; DB 1; Length 793;
 Best Local Similarity 22.9%; Pred. No. 5e-19;
 Matches 114; Conservative 62; Mismatches 149; Indels 173; Gaps 17;
 QY 8 GVATFVLQEE-FDRISGVYWCPEAETTPGGKILRLYBENDESEVEIIV----- 57
 Db 214 GLTDWYEEVSVSSPTIWWSPDS-----KIAFKLNSEIPTYHYPIYTAELDP 264
 QY 58 TSPMLETRADSFRTKGTANPKYTFKSEI----- 89
 Db 265 SLPEPDYNKMAIKIPKPNPVSLSFADLNSNASSNFSIMHNEPLAPVQVNLVWN 324
 QY 90 -----MIDAEGRIL----- 98
 Db 325 TSSVLVQFTNRNSTCITARLDLTDELKSIHTVTECLBEGWYVQSAKAPFLNLSLVWEN 384
 QY 99 -----VDEVRRL-----VPEEGTKDSP 115
 Db 385 WSDGYFDLILADDDYNNHIAFIPNGSSPIYIATNSGADVDTGPIHIDGFGNVTALATIKDS 444
 QY 116 LEHNLVYVSVNGEVTRLTDGRYS---HSCICSHODPFIKSYNQKPHCVSLYKLS 172
 Db 445 TERHLYYS-LDTLHLYGTTDGEDEGYSISFSFGDYVLYNHGPDV-----WQELKS 499
 QY 173 PEDDPTCKTEPMATILDSAGPLDPTPEI-----FSFESTGFTLYGMLYKPHDLOPK 228
 Db 500 TKDQXCYSLSTNSLKKQ-----LSSITLPSVEYGLTNDTL-FNF--MERPRPNVDVK 553
 QY 229 KYPITVLFYGGPOVQVUNNRFKGVYFRNLATLSIG-----YVVVVIDNRSGCHRGKPF 282
 Db 554 KIPVYFAYVGGGSOV-----ATLFFVNDPQAYLASHPDEFFIVTLIDRGRTGFGNAF 607
 QY 283 EGAFYKMGQLEIDOVGLQYLAIRYFIDLDRIHGWSYGGYLSLMAIMQSRDIFRY 342
 Db 608 RYSVSRHIGWESIDQAGAGFMAD-LPFDVNHVIGWMSGYGLTKTL-ETQDVFSY 665
 QY 343 AIAGAPVTLMLFYDTGYT 360

Db 666 GMAVAPYDWRLYDSVYT 683

RESULT 5
DPP4_MOUSE STANDARD; PRT; 760 AA.
AC P28843;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (Thymocyte-activating molecule) (THAM).
GN DPP4 OR CD26.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Thymus;
RX MEDLINE=92129288; PubMed=1370813;
RA Marguet D.A., Bernard A.-M., Vivier I., Darmoul D., Naquet P., Pierres M.;
RT "cDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a subgroup of serine proteases.";
RT J. Biol. Chem. 267:2200-2206 (1992).
RN [2]
RP REVISIONS.
RA Marguet D.A.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.A; TISSUE=Liver;
RX MEDLINE=95092780; PubMed=7999781;
RA Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;
RT "Structure of the mouse dipeptidyl peptidase IV (CD26) gene.";
RT Biochemistry 33:15204-15214 (1994).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Brownstein M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butcherfield V.S.N., Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [5]
RP SEQUENCE OF 1-20.
RX MEDLINE=91302787; PubMed=1712807;
RA Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y., Bernard A.-M., Gorvel J.-P., Pierres M.;
RT "Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptidase IV).";
RT J. Immunol. 147:447-454 (1991).
CC -I- IMMUNOL: 147:447-454 (1991).
CC polypeptides having N-terminal dipeptides sequentially from penultimate residue is proline.

CC -I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb|-Xcc, from a polypeptide, preferentially when Xbb is Pro, provided Xcc is neither Pro nor hydroxyproline.
CC -I- SUBUNIT: Homodimer.
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in a soluble form.
CC -I- PTM: The soluble form (SDP) derives from the membrane form (MDP) by proteolytic processing.
CC -I- SIMILARITY: Belongs to peptidase family S9B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58384; CAA41274.1; -
DR EMBL: U12620; AAA82213.1; -
DR EMBL: U12599; AAA82213.1; JOINED.
DR EMBL: U12600; AAA82213.1; JOINED.
DR EMBL: U12601; AAA82213.1; JOINED.
DR EMBL: U12602; AAA82213.1; JOINED.
DR EMBL: U12603; AAA82213.1; JOINED.
DR EMBL: U12604; AAA82213.1; JOINED.
DR EMBL: U12605; AAA82213.1; JOINED.
DR EMBL: U12606; AAA82213.1; JOINED.
DR EMBL: U12607; AAA82213.1; JOINED.
DR EMBL: U12608; AAA82213.1; JOINED.
DR EMBL: U12609; AAA82213.1; JOINED.
DR EMBL: U12610; AAA82213.1; JOINED.
DR EMBL: U12611; AAA82213.1; JOINED.
DR EMBL: U12612; AAA82213.1; JOINED.
DR EMBL: U12613; AAA82213.1; JOINED.
DR EMBL: U12614; AAA82213.1; JOINED.
DR EMBL: U12615; AAA82213.1; JOINED.
DR EMBL: U12616; AAA82213.1; JOINED.
DR EMBL: U12617; AAA82213.1; JOINED.
DR EMBL: U12618; AAA82213.1; JOINED.
DR EMBL: BC022183; AAH22183.1; -
DR EMBL: BC022183; AAH22183.1; -
DR MGD: MGD:94919; DPP4.
DR InterPro: IPR002459; DPP4 N-term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR000379; PIOL_endopep_ser.
DR Pfam: PF00930; DPP4V N-term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 760
FT DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT (MDP).
FT CHAIN 37 760
FT DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT (SDP).
FT DOMAIN 1 6
FT CITOPLASMIC (POTENTIAL).
FT TRANSMEM 7 28
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 29 760
FT CYC-RICH.
FT ACT_SITE 624 624
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 702 702
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 83 83
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 113 113
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 213 213
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 223
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 315 315
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 328 328
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 514 514
FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CONFLICT	332	332	T -> N (IN REF. 2)	
FT	CONFLICT	352	352	T -> V (IN REF. 2)	
FT	CONFLICT	394	394	V -> D (IN REF. 2)	
FT	CONFLICT	562	562	L -> F (IN REF. 2)	
FT	CONFLICT	624	624	R -> Q (IN REF. 2)	
SO	SEQUENCE	767 AA;	88003 MW;	AA17COEC6F054652 CRC64;	
Query Match 15.4%; Score 297.5; DB 1; Length 767;					
Best Local Similarity 24.5%; Pred. No. 9.4e-18;					
Matches 121; Conservative 60; Mismatches 152; Indels 161; Gaps 22					
QY	8	GVAITYLQGE-FDRSGVWMCCKATTTSQSGKILRIIYENDESEVEIHHVTSPLMLTRR	66		
DB	195	GINDWVYEELIEIGASALMW-----SPNGFIAYAGFNDGVPLIEFSYSSESICQYK	248		
QY	67	ADSFRRPKGTANPKVTF-----			
DB	249	TWMIPIPKAGAVNPTKPIFIVNTDLSSTTTIMQITAPASVTTGPDHLCDAVMSSEDR	308		
QY	97	IIVDERRLVTF-----EGTDS-----PLEHLLVVSIV	126		
DB	309	ISIQMLRKIQNSVMAICDYDKTILWMNCPTTQHIETISATGCGRRPAPRH-----FT	363		
QY	127	NPGE-VTILTR-GYSHSCCI-----SQHCFD-----FISKYSNQ-	159		
DB	364	SDGSSPYKIVSDKDGKHLCCQKQKREQVCTFTGAMAVISIALTSIDLYIISNEY	423		
QY	160	KMPHCVSIVK-----LSPEDDPCK-----IKEMATILISAGL-LPDTY-	199		
DB	424	KEMPGGRNLIKQLDHTNKKLSCDLPERCQYYSVLSKKAQYVOLGCGGGLPELYTL	483		
QY	200	-----PPEIFSFESTGTFT-----LYGMLYKPHLDQPKKP	231		
DB	484	HRSTQDEKRLVEDNSALDKMLQDVMPSKKIDFIYENETRFVYQMLIRPH-FDSKXKP	542		
QY	232	TVLFIYGGQVQVLYNNRFGVYFFELN--TLASLGIVVVV-IDNGSCHRLKPEGARK	287		
DB	543	LILDIVAGPQSQ-----KADAFKFLNATYTLASTENIIVASFGRSSGQGGDKIMHAIN	596		
QY	288	YKMGQLEIDDQVEGL-QYLASRYDFIDIDRYGHHGWSYGYLSTLMALMORSIDIFRVAIAG	346		
DB	597	KRLGTLEVEDQLEAARQFL--KMGFVDSKRFVAILMGWSYGGVYVSMVLGSGVFKGGLIAY	654		
QY	347	APYTLWIFPYDTGTY	360		
DB	655	APVSRWEYDYDSVT	668		
RESULT 7					
DAP2_YEAST					
ID	DAP2_YEAST	STANDARD;	PRT;	818 AA.	
AC	P18962;				
DT	01-NOV-1990 (Rel. 16, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DPAp B) (YSCV).				
GN	DAP2 OR YHR028C.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
OX	NCBI_TaxId=4932;				
RN	[1]				
RX	SEQUENCE FROM N.A.				
RA	MEDLINE=89174971; PubMed=2647766;				
RA	Roberts C.J., Pohlig G., Rothman J.R., Stevens T.H.;				
RT	"Structure, biosynthesis, and localization of dipeptidyl				
RT	aminopeptidase B, an integral membrane glycoprotein of the yeast				
RL	vacuole.";				
RL	J. Cell Biol. 108:1363-1373(1989).				
RP	[2]				
RC	SEQUENCE FROM N.A.				
RC	STRAIN=S288C / AB972;				
RC	MEDLINE=94378003; PubMed=8091229;				

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favallo A., Fulton L., Gattung S., Geisel C., Kresen J.,
 RA Kuchaba T., Hillier L., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
 RA Nman M., Rifkin L., Riles L., St Peter H., Trevasaks E., Vaughan K.,
 RA Vignati D., Wilcox L., Woldman P., Waterston R., Wilson R.,
 RA Vaudin M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 VIII." ;
 RL Science 265:2077-2082(1994).
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
 CC VACUOLES.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; X15484; CAA33512.1; -;
 CC DR EMBL; U10399; AAB68879.1; -;
 CC DR PIR; S46780; A30107.
 CC DR MEROPS; S09.006; -;
 CC DR SGD; S00001070; DAP2.
 CC DR GO; GO:0000329; Cyvacuolar membrane (sensu Fungi); IDA.
 CC DR Interpro; IPR002469; DEPIV_N term.
 CC DR Interpro; IPR001375; Peptidase_S9.
 CC DR Interpro; IPR0002471; Prol_endopep_ser.
 CC DR Interpro; IPR0003479; Ser_sertr_site.
 CC DR Pfam; PF00930; DEPIV_N term; 1.
 CC DR Pfam; PF00326; Peptidase_S9; 1.
 CC DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 CC DR Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
 CC KW Transmembrane; Glycoprotein; Signal-anchor.
 CC KM CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 1 29
 CC FT TRANSSEM 30 45
 CC FT
 CC FT DOMAIN 46 818
 CC FT ACT_SITE 679 679
 CC FT ACT_SITE 756 756
 CC FT ACT_SITE 789 789
 CC FT CARBOHYD 63 63
 CC FT CARBOHYD 79 79
 CC FT CARBOHYD 110 110
 CC FT CARBOHYD 139 139
 CC FT CARBOHYD 372 372
 CC FT CARBOHYD 392 392
 CC FT CARBOHYD 421 421
 CC FT CARBOHYD 738 738
 CC FT CONFLICT 83 83
 CC FT CONFLICT 125 125
 CC FT CONFLICT 182 188
 CC FT CONFLICT 200 200
 CC FT CONFLICT 366 375
 CC FT CONFLICT 808 818
 CC FT
 CC SQ SEQUENCE 818 AA; 93404 MW; 318F50445375BD3 CRC64;
 QY Query Match 14.5%; Score 280.5; DB 1; Length 818;
 Db Best Local Similarity 22.0%; Pred. No. 3e-16;
 QY Matches 114; Conservative 50; Mismatches 155; Indels 199; Gaps 19;
 QY 8 GVATPVLQEE-DRDRSGYWMCPKALETTPSGGKILIRIYBENDESEY-ELIHVTSPLETR 65
 Db 234 GRPDWYBEVEVEDDKAMWSPGTD-----YLAFLKIDSEVGERF---IPYYVD 281
 QY 66 RAD-----SFRYPPTGTAND----- 80
 Db 282 EKDIYBEMRSIKYPSKSGTPNPHALWVYSMKDGTSPHPRIKSGNKKDGSLLITREVTWVGNG 341

```

QY 81 -----KTFKMSI-----MID-----92
DB 342 NLVKTTRSSDILITVLEIOTIAKTSNVNNESSNGMWETHTNLEFIPANETPRPHNG 401
QY 93 -----AEGRIIVDEVRRLVYEGTQDPLH 118
DB 402 YVDILPFIGYNHLAFENSSSHKKTTEGKMEVNVNPLAFDSKENVRLYFISTKRSSTER 461
QY 119 HLTVSVYVNGEVTRLTDRGYS--HSCCISQHCDFPISKYSNOKNPH-----163
DB 462 HYYIIDLRSPEIIEVTDTSDDGYDVSFSSGRFGLITLKGPVVPQKIVDHSRAEK 521
QY 164 C-----VSLYLTSSDEDPCTCKTFMATTILDSAGPLPDYLPPELIFSESTGTGLYG 216
DB 522 CDKNVVGKSLYHLEKNE-----VLTKE--LEDYAVPRKSFRELMLGKDEFG 566
QY 217 --MLYKPHDLP-----GKKYPTVLFYIGGPOVOLVNNRFGVKKERL-----TLASL 263
DB 567 KQILVMSYELLPDPFDETLSDHVPFFAYGGRNSQV-----VKTFSGFNEVASQL 620
QY 264 GYVVVVVINDNGSCHRGKKEFGAKFYKMGQIEIDQVEGLQYLAARYDFIDLDVGHGWS 323
DB 621 NALVVVVVDGRTGTFKQDFRSLVRLDGYEARDOISAASLYGS-LTFVDPQKISLFGWS 679
QY 324 YGGYLSLMLMQRSD-IFRYAIAAGAPVTLMIPTDTGTT 360
DB 680 YGGYLTTLTKTEKQGRHFKMGSVAPVDMRFDSVYT 717

RESULT 8
SEPR_MOUSE
ID SEPR_MOUSE STANDARD; PRT; 761 AA.
AC P97321;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Seprese (BC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
  membrane serine protease).
GN PAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1, 2 AND 3).
RC STRAIN=BALB/c; TISSUE=Embryo;
RX MEDLINE=97284459; PubMed=9139873;
RA Klausner R.D., Scandian M.J., Garin-Chesa P., Dalber C., Fiebig H.H.,
RA Old L.J., Rettig W.J., Schnapp A.;
RT "Mouse fibroblast activation protein: molecular cloning, alternative
RT splicing and expression in the reactive stroma of epithelial
RT cancer.";
RT Int. J. Cancer 71:383-389(1997).
RL [2]
RL SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=Breast;
RX MEDLINE=22388257; PubMed=12477932;
RA Strussberg R.L., Feigold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buettow K.H., Scheffer E., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Stapleton M., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.O., Usdin T.B., Caesvart T.L., Scheetz T.E.,
RA Rana S.S., Loquellain N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bolek S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Mizny D.M., Sodergren E.U., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blaesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez J.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,

```

```

RA Schnorch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: May have a role in tissue remodeling during development
CC and wound healing, and contribute to invasiveness in malignant
CC cancers.
CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC type IV collagen, but not native type I or type IV collagen. Does
CC not cleave laminin, fibronectin, fibrin or casein.
CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC inactive (by similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (by similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Name=1;
CC IsoId=P97321-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P97321-2; Sequence=VSP_005368;
CC Name=3;
CC IsoId=P97321-3; Sequence=VSP_005369;
CC -1- TISSUE SPECIFICITY: Detected in fibroblasts, in placenta, uterus,
CC embryos from day 7-19 and in new-born mice (p1).
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y10007; CAJ1116.1; --
DR EMBL: BC019190; AAH19190.1; --
DR MEROPS: S09.007; --
DR MGD; MGI:109608; Pap.
DR InterPro: IPR002469; DPPIV N term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR002471; Pro_endopep_ser.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV N term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
DR Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;
KM Glycoprotein; Alternative splicing.
FT DOMAIN 1 4
FT TRANSMEM 5 25
FT FT 4
FT FT 4
FT DOMAIN 26 761
FT ACT_SITE 624 624
FT ACT_SITE 702 702
FT ACT_SITE 734 734
FT CARBOHYD 49 49
FT CARBOHYD 92 92
FT CARBOHYD 99 99
FT CARBOHYD 227 227
FT CARBOHYD 314 314
FT CARBOHYD 679 679
FT VARSPLIC 31 35
FT VARSPLIC 31 35
FT VARSPLIC 63 63
FT VARSPLIC 737 737
FT CONFLICT 737 737
FT SEQUENCE 761 AA; 87944 MW; 914C3AED8213B25 CRC64;
SQ
Query Match 13.4%; Score 259; DB 1; Length 761;
Best Local Similarity 21.2%; Pred. No. 2e-14;
Matches 106; Conservative 66; Mismatches 149; Indels 178; Gaps 18;
QY 8 GYATVYLOEER-DRYSYWCPCRAETTPSGGKIIRIYENDESEVEIHW-----S 59
DB 195 GIPDVYEEEMLATKYA-LMWSP-----DGKFL--AVYEPNDSDIPILAVSYGDQGY 244

```


QY 60 PMLETRRADSPRYKGTGNPKY-----TFK 85
 Db 245 P-----RTINIPYKAGAKNEVAVFIVDTTYPHHVGPMEVPEMTASSDYFSWLTWV 299
 QY 86 MSE----- 88
 Db 300 SSEKVCLOMLKRVGNVSLICDPREDHNAMECPKNOHVEESTGTGAGCFVSTPAFSQ 359
 QY 89 -----IMDAEG-----RIIVDEVRIV-----YFEGTKDS-----P 115
 Db 360 DATSYVKIFSDKDGKHIHYIKDVTENAIQITSGKEWAIYFRTVQSLFYSSNEFGYF 419
 QY 116 LEHLIYVSYNPEGEVTRLDRGYSHSCISQH-----CPEFISKYNGNCPHCVSLY-- 168
 Db 420 GRNRIYKISGNPSPSK-----CYTCHLRKRCQYTAFSFYKAKYALVCGP 469
 QY 169 --KLSSPEDDPTCKEKFMAITLDSAGPLDPYTPPEI--FSFESTTGTLYGMLYKPHDLQ 225
 Db 470 GLPISTLHDKRTDDEIQVLEENKELNSLNICLPKVEIKKDKGGLTFYKMLLPQFD 529
 QY 226 PGKKYPTVFLTYGGPQVLYNNRKGVKRYRLNTLASL---GYVVVINDRSGCHRGK 281
 Db 530 RSKKYPPLIQLVYGGPCQSQSVKS-----VPAVNWITYLAKSEKIGIVIALVDGRGTAFGDK 583
 QY 282 FEAGFKYKMGQIEIDDOVEGLOYLASRYDFIDLRVGIHGMVGYGLTMAIMQRSDIER 341
 Db 584 FLNAVYKGLGVYEVEDQLTAVKRTI-EMGFIDEIRAIWMSYGVSLALASGTLFPK 642
 QY 342 VAIAGAPVTLWTFYDTGYT 360
 Db 643 CGIAVAVPSSMEYASIYS 661

RESULT 9
 SEPR_HUMAN
 ID SEPR_HUMAN STANDARD; PRT; 760 AA.
 AC Q12884; Q00199; Q09998; Q0UID4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
 DE membrane serine protease) (170-kDa melanoma membrane-bound
 DE gelatinase).
 GN FAP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_Taxid=9606;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Fibroblast;
 RX MEDLINE=94261645; PubMed=7911242;
 RA Scanlan M.J., Raj B.K.M., Celivo B., Garin-Chesa P., Sanz-Noncasi M.P.,
 RA Healey J.H., Old L.J., Rettig W.J.;
 RT "Molecular cloning of fibroblast activation protein alpha, a member of
 RT the serine protease family selectively expressed in stromal
 RT fibroblasts of epithelial cancers";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:5657-5661(1994).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Melanoma;
 RX MEDLINE=97388251; PubMed=9247085;
 RA Goldstein L.A., Ghersi G., Pineiro-Sanchez M.L., Salame M., Yeh Y.,
 RA Flesate D., Chen W.-T.;
 RT "Molecular cloning of seprase: a serine integral membrane protease
 RT from human melanoma";
 RL Biochim. Biophys. Acta 1361:11-19(1997).
 RN (3)
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 220-229; 461-472 AND
 RP 511-518.
 RC TISSUE=Melanoma;
 RX MEDLINE=97218181; PubMed=9065413;

RA Pineiro-Sanchez M.L., Goldstein L.A., Dodt J., Howard L., Yeh Y.,
 RA Chen W.-T.;
 RT "Identification of the 170-kDa melanoma membrane-bound gelatinase
 RT (seprase) as a serine integral membrane protease";
 RL J. Biol. Chem. 272:7595-7601(1997).
 RN (4)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Melanoma;
 RX MEDLINE=20112818; PubMed=10644713;
 RA Goldstein L.A., Chen W.-T.;
 RT "Identification of an alternatively spliced seprase mRNA that encodes
 RT a novel intracellular isoform";
 RL J. Biol. Chem. 275:2554-2559(2000).
 RN (5)
 RP SEQUENCE OF 192-208; 220-240 AND 510-521.
 RX MEDLINE=94327249; PubMed=7519584;
 RA Rettig W.J., Su S.L., Fortunato S.R., Scanlan M.J., Raj B.K.M.,
 RA Garin-Chesa P., Healey J.H., Old L.J.;
 RT "Fibroblast activation protein: purification, epitope mapping and
 RT induction by growth factors";
 RL Int. J. Cancer 58:385-392(1994).
 CC -1- FUNCTION: May have a role in tissue remodeling during development
 CC and wound healing, and may contribute to invasiveness in malignant
 CC cancers.
 CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
 CC type IV collagen, but not native type I or type IV collagen. Does
 CC not cleave laminin, fibronectin, fibrin or casein.
 CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
 CC inactive.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Found in cell
 CC surface lamellipodia, invadopodia and on shed vesicles.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=L;
 CC IsoId=Q12884-1; Sequence=Displayed;
 CC Note=Major isoform;
 CC Name=2; Synonyms=S, Truncated;
 CC IsoId=Q12884-2; Sequence=VSP_005367;
 CC -1- TISSUE SPECIFICITY: Fibroblast-specific.
 CC -1- INDUCTION: In fibroblasts at times and sites of tissue remodeling
 CC during development, tissue repair, and carcinogenesis.
 CC -1- PTM: N-glycosylated.
 CC -1- PTM: The N-terminus may be blocked.
 CC -1- SIMILARITY: Belongs to peptidase family 99B.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U09278; AAB94652.1; -
 CC EMBL; U76833; AAC51668.1; -
 CC EMBL; AF007822; AAF21600.1; -
 CC MEROPS; S09.007; -
 CC GeneW; HGNC:3590; FAP.
 CC MIM; 600403; -
 CC InterPro; IPR002469; DDPV N term.
 CC InterPro; IPR001375; Peptidase_S9.
 CC InterPro; IPR002471; Pro_endopep_ser.
 CC InterPro; IPR000379; Ser_estrs_site.
 CC Pfam; PF00930; DDPV_N term; 1.
 CC Pfam; PF00326; Peptidase_S9; 1.
 CC PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 CC HydroLase; Protease; Serine protease; Transmembrane; Signal-anchor;
 CC Glycoprotein; Alternative splicing;
 CC DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC FT DOMAIN 26 760 (POTENTIAL).
 CC FT ACT_SITE 624 624 EXTRACELLULAR (POTENTIAL).
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 702 702 CHARGE RELAY SYSTEM (BY SIMILARITY)
 FT ACT_SITE 734 734 CHARGE RELAY SYSTEM (BY SIMILARITY)
 FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPIC 1 521 Missing (in isoform 2).
 FT CONFLICT 207 207 /FTid=VSP_005367.
 FT CONFLICT 229 229 P -> A (IN REF. 2).
 FT CONFLICT 354 354 K -> T (IN REF. 2).
 FT CONFLICT 354 354 R -> T (IN REF. 2).
 SQ SEQUENCE 760 AA; 87820 MM; A0D34B4801B07BA CRC64;

Query Match 13.3%; Score 257; DB 1; Length 760;
 Best Local Similarity 22.7%; Pred. No. 2,9e-14;
 Matches 110; Conservative 68; Mismatches 158; Indels 148; Gaps 20;

8 GVATFLOEEF--DRYSGVWMPCKAETPGSGKILRLIYENDESEVEIHTVSPMLET- 64
 195 GIDWVYEEEMLETKIA-LWMSP-----NGKFL--AAAEFNDKDIPIVLAISYGDDEQY 244
 65 RRADSEFPYKVTGTANPKVTFKM-----SEIMIDA-----EGRI 97
 245 PRINIPYPRGAKNPVVRIFIDTTPAVVGQEVVPVPMIASSDYPSMLTWIDERY 304
 98 IYDEVRALVYFE-----GTQDSP-LEHLY-----VVSYPGE 130
 305 CLQMLKRVQNVSVLSICDFREDWQTDCEPTQHEIESRTGMAGCFVSPRPVSYSALISY 364
 131 VTRLTDR-GYSHSCCI-----SQHDFEIS---KYSQKRP 162
 365 YKIFSDGKGIKIHVKQIVENALQITSGKMEINIFRYVQDSLFSSNEFEERYPRRNI 424
 163 HCVSLYKLSPPEDDPTCKTE---FMATILDSA-----GP-LPYT----- 199
 425 YRISIGSYPSPKCVCHLRKRCQYRTAFSPYAKYVALVCGPEIPSTLHDKRTDGE 484
 200 -----PPRIFSESTTGFTLYMLYKPHDLQPKKYPTVLFYGGP 240
 485 IKILENKELENALKNIOLEPKBEIKKLEVEDEITLWYKMLIPQDPDRSKKYPILLIYVGGP 544
 241 QVALNNRPFKGVYFRLNLTASL---GYVVVVINDRSGCHGKKEGAFKRMQIETD 296
 545 CSQSR-----VRAVMISTYLASKEGVNIALVDRGTAFCGDKLIVAVYKRLGYEYE 598
 297 DVEGLQYLASYRDFIDRVGIGHWSYGYLSLMAIMQSRDIFRVAIAGAPVTLMIFYD 356
 599 DQITAVRKPI-EMGFIDEKRIAIWMSYGYVSSIALAGTGLFCGIALVAIVSSMEYVA 657
 357 TGYT 360
 658 SVYT 661

RESULT 10
 STL3_YEAST STANDARD; PRT; 931 AA.
 AC P33894;
 DT 01-FEB-1994 (Rel. 28; Created)
 DT 28-FEB-1994 (Rel. 28; Last sequence update)
 DT 28-FEB-2003 (Rel. 41; Last annotation update)
 DE Dipeptidyl aminopeptidase A (PC 3.4.14.1-) (DPAP A) (YSCIV).
 GN STL3 OR YCII OR YOR219C OR YOR50-9.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9506382; PubMed=7975897;
 RA Anna-Miriola S.S., Herskowitz I.;
 RT "Isolation and DNA sequence of the STL3 gene encoding dipeptidyl

RT aminopeptidase.";
 RL Yeast 10:801-810 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AB320;
 RA Flanagan C.A., Thorne J.;
 RT "STL3";
 RL (in) Getting M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.);
 RL Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
 Press, Oxford (1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=96437977; PubMed=8840505;
 RA Galisson F., Dujon B.;
 RT "Sequence and analysis of a 33 kb fragment from the right arm of
 chromosome XV of the yeast *Saccharomyces cerevisiae*.";
 RL Yeast 12:877-885 (1996).
 CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
 ALPHA-FACTOR PRECURSOR.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
 VACUOLES.
 CC -1- SIMILARITY: Belongs to peptidase family 59B.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; L21944; AAA5119.1; -
 EMBL; U08230; AAA17897.1; -
 EMBL; X92441; CAA53182.1; -
 EMBL; Y5127; CAA9437.1; -
 FIR; A49737; A49737.
 DR MEROPS; S09.005; -
 DR SGD; S0005745; STR13.
 DR GO; GO:0005802; C:Golgi trans-face; IDA.
 DR GO; GO:0004177; P:aminopeptidase activity; IDA.
 DR GO; GO:0007323; P:peptide pheromone maturation; IDA.
 DR InterPro; IPR002469; D:Peptid N term.
 DR InterPro; IPR001375; Peptidase S9.
 DR InterPro; IPR002471; Prol endopep. ser.
 DR InterPro; IPR00379; Ser_cstrs_site.
 DR Pfam; PF00330; DPEPTV_N_term; 1.
 DR Pfam; PF00326; Peptidase S9; 1.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; FALSE NEG.
 KW Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;
 KW Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.
 FT DOMAIN 1 119
 FT TRANSMEM 120 140
 FT DOMAIN 141 931
 FT ACT_SITE 785 785 LUMENAL (POTENTIAL).
 FT ACT_SITE 863 863 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 896 896 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 931 AA; 107200 MM; 81AF70094093C023 CRC64;

Query Match 12.9%; Score 250; DB 1; Length 931;
 Best Local Similarity 24.3%; Pred. No. 1.5e-13;
 Matches 101; Conservative 58; Mismatches 162; Indels 94; Gaps 15;

15 QEEPRYSYGMWCP---KAEPTPGSKILRLIYENDESEVEIHTVTS----- 59
 434 QKDSILYNGKMWISPDTPFRFEITDRNSKILDVAVYDIPSSQMLTVANTSNLFGNIETKX 493
 60 -----PVLSTR-----ADS-----FRYKGTGTANPKYFKKSEIMIDAEGRI 98
 494 DLSIPPELAKMDYGIYDIHADSRRGSHLYFP-TVFAKEPIQLTKGNWEVTGNGIVG 552

QY 99 VDEVRRLVYFEGTKDPSLEHLLVYVSYV-----NPGVETRLTDRGVSHCSCIS 146
 Db 553 YEVETDITFTFANISIGMSQHLVLSILDTSTQNTQPSLQNP-----SDKYDPIDELS 606
 QY 147 QHCDFFISKYSNQNP-----HCVSLYKLSSPEDPTCKTKEMWATILD 190
 Db 607 SSARVAISKXGLGPDTPKIVAGPLTRVLNVAEIHDDSLILQTLKDE-----KFKK----- 654
 QY 191 SAGLPDPTPELFEFESTT---GFTLYGMLYKPHDLOPKKPTVTLFTYGGPOVLVNN 247
 Db 655 ---KIKNYDLP-LTSYTWVLDDGVEINYLEIKRANINPKKPYILVNIYGGSGQTFTT 710
 QY 248 RFGVKYFRINTLASLGYVVVVIDNRGSGRGLKFECAFYKKGQLEIDQVE-GIQYLA 306
 Db 711 --KSLAFEGAVVSGLDVILQLEPRGTGKGMSFRSMAEKLYGMEPRDITEVTKKFI 768
 QY 307 SRDPTDLDRVGHGMSYGYLSLML-MQRSDIFRVAIAGAPVTLWTFYDTGYT 360
 Db 769 RNSQHIDESKIAIWMSYGGFTSLKTVELNDGDTFKYMAVAAPVMTWLTLYDSVT 823

RESULT 11

DPPE BOVIN STANDARD; PRT; 863 AA.

ID DPPE BOVIN STANDARD; PRT; 863 AA.
 AC P42659;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
 DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
 GN DPP6.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 NC NCB1_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S), AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=92108018; PubMed=1729689;
 RA Wada K., Yokocani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
 RT "Differential expression of two distinct forms of mRNA encoding
 RT members of a dipeptidyl aminopeptidase family.";
 RT Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
 CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=DPPX-L;
 CC IsoId=P42659-1; Sequence=Displayed;
 CC Name=DPPX-S;
 CC IsoId=P42659-2; Sequence=VSP 005364;
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L
 CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN
 CC BRAIN, KIDNEY, OVARY AND TESTIS.
 CC -!- SIMILARITY: Belongs to peptidase family 89B.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; M76428; AAC41622.1; -;
 CC EMBL; M76429; AAC41623.1; -;
 CC PIR; A41793; A41793.
 CC MEROPS; S09.973; -.

DR InterPro; IPR002469; DPPV N term.
 DR InterPro; IPR001375; Peptidase 59.
 DR InterPro; IPR000379; Ser esters site.
 DR Pfam; PF00930; DPPV N term; 1.
 DR Pfam; PF00326; Peptidase S9; 1.
 KW Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
 FT DOMAIN 1 93
 FT TRANSMEM 94 114
 FT 115 863
 FT DOMAIN 171 171
 FT CARBOHYD 402 402
 FT CARBOHYD 469 469
 FT CARBOHYD 533 533
 FT CARBOHYD 564 564
 FT CARBOHYD 811 811
 FT VARSLIC 1 79
 FT FT
 SQ SEQUENCE 863 AA; 96556 MW; 23DBA792B841A39D CRC64;
 Query Match 11.4%; Score 221; DB 1; Length 863;
 Best Local Similarity 21.0%; Pred. No. 4, Se-11;
 Matches 103; Conservative 58; Mismatches 169; Indels 160; Gaps 16;

QY 8 GVATFVLOAE-FDRYSGYWCCKAETTPGKILRLIYENDESEVEIIVH--TSPLME 63
 Db 285 GSDMLYEELIKTHIAHMSPDG-----TRLAATINDSRVWELPYTOSVYP 335
 QY 64 TRADSFRKPKGTANPKVTF-----KXSEIMI----- 91
 Db 336 T-AKRYHPKACGNCNPISLHVIGLNGPTHLEMTPPDDPRRREYITWVKWATSTKVA 393
 QY 92 -----DA-----EGR----- 96
 Db 394 VNMISRAQNVSLITLCDATGTGCTKHEDSEAMLIHRONEBPVFSKDGKFFVRRAIPQG 453
 QY 97 -----IYDEVRRLVYFEGTKDPSLEHLLY 121
 Db 454 GQGRFYHTVSSSQPNSSNDNIQSITSGDMVDTKLISYDKRSQYIFLTBDLPRRQLY 513
 QY 122 VSYVNPGEVTR-----LTDRGYHSCCISQHCPPFSIKYSNQNPCHVSLYKSSPED 175
 Db 514 SASVY--GSFNNQCLSCDLVNDCTVFSASFSGADEFLLKCBGPVP-TVSVANTDKKK 570
 QY 176 DPTCKTKE-FWATILDSAGPLDYPPPELFEFESTTGTLYGMLYKPHDLOPKKPTVTL 234
 Db 571 MEDLETNEHVQKALSDROMPKVEYRKIE-----TDYKPLPIQLKPATFTDTAHPYLL 624
 QY 235 FIYGGPOVLVNNRFGYKFRLLNTLASLGYVVVVIDNRGSGRGLKFECAFYKKGQLE 294
 Db 625 VVDGTGPGSGVAEKE--AVTWETVWVSSHGAVVCDGRSGFGQTRLHLVERRRLLSLE 682
 QY 295 IDDOVEGQIYASLRDPTDLDRVGHGMSYGYLSLMLMQRSD---IFRVAIAGAPV 350
 Db 683 EKDQWEAVRWAL-KEPYIDKTRVAVFGDYGYSYLLIPAKGDAQAPVFCSGSALSPIT 741
 QY 351 LWFYDTGYT 360
 Db 742 DFKLYASAFS 751
 RESULT 12
 DPPE RAT STANDARD; PRT; 859 AA.
 ID DPPE RAT STANDARD; PRT; 859 AA.
 AC P46101;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
 DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
 GN DPP6.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 NC NCB1_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S), AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=92108018; PubMed=1729689;
 RA Wada K., Yokocani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
 RT "Differential expression of two distinct forms of mRNA encoding
 RT members of a dipeptidyl aminopeptidase family.";
 RT Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
 CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=DPPX-L;
 CC IsoId=P42659-1; Sequence=Displayed;
 CC Name=DPPX-S;
 CC IsoId=P42659-2; Sequence=VSP 005364;
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L
 CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN
 CC BRAIN, KIDNEY, OVARY AND TESTIS.
 CC -!- SIMILARITY: Belongs to peptidase family 89B.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; M76428; AAC41622.1; -;
 CC EMBL; M76429; AAC41623.1; -;
 CC PIR; A41793; A41793.
 CC MEROPS; S09.973; -.

GN DPP6.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN NCBI_TaxID=10116;
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
 RC TISSUE=Brain.
 RX MEDLINE=92108018; PubMed=1729689;
 RA Wada K., Yokotani N., Hunter C., Doi K., Wentholt R.D., Shimasaki S.;
 RT "Differential expression of two distinct forms of mRNA encoding
 RT members of a dipeptidyl aminopeptidase family";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
 CC -1- FUNCTION: May be involved in the physiological processes of brain
 CC function. Has no dipeptidyl aminopeptidase activity. The lack of
 CC activity may be due to the substitution of an aspartate residue
 CC for the serine residue in the proposed catalytic triad.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=DPPX-L;
 CC IsoId=P46101-1; Sequence=Displayed;
 CC Name=DPPX-S;
 CC IsoId=P46101-2; Sequence=VSP 005366;
 CC -1- TISSUE SPECIFICITY: DPPX-S IS EXPRESSED IN BRAIN AND SOME
 CC PERIPHERAL TISSUES INCLUDING KIDNEY, OVARY, AND TESTIS; IN
 CC CONTRAST DPPX-L IS EXPRESSED ALMOST EXCLUSIVELY IN BRAIN.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M76426; AAC42061.1; -;
 DR EMBL: M76427; AAC42062.1; -;
 DR MEROPS: S09.973; -;
 DR InterPro: IPR002469; DPPV N term.
 DR InterPro: IPR001375; Peptidase S9.
 DR InterPro: IPR000379; Ser estersite.
 DR Pfam: PF000930; DPPV N term; 1.
 DR Pfam: PF00326; Peptidase S9; 1.
 KM Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
 FT DOMAIN 1 89
 FT TRANSMEM 90 110
 FT SIGNAL-ANCHOR (POTENTIAL).
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 111 859
 FT CARBOHYD 167 167
 FT CARBOHYD 168 168
 FT CARBOHYD 313 313
 FT CARBOHYD 398 398
 FT CARBOHYD 465 465
 FT CARBOHYD 529 529
 FT CARBOHYD 560 560
 FT CARBOHYD 807 807
 FT VARSPLIC 1 75
 FT EXTRACELLULAR (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MASLYRFTGKINTSSFPAPPAASHLGGQGEEDAGSKP
 FT LGPOADVAAPERAGAGRRFYOARSDCEDD -> MTT
 FT AKPSASGSKVQGOQD (in isoform DPPX-S).
 FT /FtId=VSP 005366.
 SQ SEQUENCE 859 AA; 97301 MW; CE26856D26ED126B CRC64;
 Query Match 11.0%; Score 212; DB 1; Length 859;
 Best Local Similarity 19.7%; Pred. No. 2.7e-10;
 Matches 98; Conservative 67; Mismatches 158; Indels 174; Gaps 16;
 Oy 8 GAATPFLQRE-FDRYGVWMPQRAETTPSGGKILRLIYENDESEVFIHV---ISPMLE 63
 Db 281 GSDMLYEEELIKSHIAHWMSPDG-----TRLAYATINDSRVPLMELPYTGSVVP 331
 Oy 64 TRRADSFYPTKTGTANPKVTF-----KMSSEIMI----- 91

Db 332 T--VKPYHYPKAGSENSISLHVLGNQPTHDLFEMPPDDPRMEEYITVWKMASTKVA 389
 Oy 92 -----DA-----BGR----- 96
 Db 390 VTWLNRAQNSILTLCDATTVCTKHEDESEAWLHRONEBPVPSKGRKFFVRAIPQG 449
 Oy 97 -----IIVDEVRRLVYEGTDSPLEHLY 121
 Db 450 GRGKFHITVSSSPNSNDNISITSGDMVDTEILYDEKANKLYFLSTDLPRRRALY 509
 Oy 122 VSYVNDGEVTRLTDRGYSHCCISQHCDFEISRYSNQKNPCHVLYKLSSP-----ED 175
 Db 510 SANIVDDENRQCL-----SCLVENCTVYASFSHMD---FFLKCEBPVPTVYH 559
 Oy 176 DPCKTEF-----WATLDSAGLPDPTPEPISFESTGTFLYGLMKYKPHDLPQ 227
 Db 560 NTDKRMFPLEANEQOKALYPMQMKIEYKRIEV-----EDYSIPMOLIKATPDT 613
 Oy 228 KKYPTVLFIVGPOVLNNRFKGVYFRLNTLASLGVVVVINDRSGCHRLKEGAFK 287
 Db 614 AHYDLIVDQGTSGSVSRFEVT--WEIVLVSHGAIVVYKCDGSGFGGTLLHEVR 671
 Oy 288 YKRGQLEIDQVBSGLQYLASRYDFIDLDRYGHGMSYGYLSLMALQKSD---IFRVA 343
 Db 622 RRLGRLERKQMEAVRMTL-KEGYIDKTRVAVFKDYGGLSTYILPAKENQGTFTCG 730
 Oy 344 IAGAPVLMIFYONGYV 360
 Db 731 SALSPITDFLYASAFS 747
 RESULT 13
 DPP6 MOUSE
 ID DPP6 MOUSE STANDARD; PRT; 804 AA.
 AC Q9Z218; Q9QW22; Q9Z219;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
 DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
 DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
 GN DPP6 OR DPP-6.
 OS Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCBI_TaxID=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Embryo;
 RX MEDLINE=99030650; PubMed=9811881;
 RA Hough R.B., Lenggeling A., Bedian V., Lo C., Bucan M.;
 RT "Rump white inversion in the mouse disrupts dipeptidyl aminopeptidase-
 RT like protein 6 and causes dysregulation of kit expression.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:13800-13805(1998).
 CC -1- FUNCTION: May be involved in the physiological processes of brain
 CC function. Has no dipeptidyl aminopeptidase activity. The lack of
 CC activity may be due to the substitution of an aspartate residue
 CC for the serine residue in the proposed catalytic triad (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF092507; AAC97366.1; -;
 DR EMBL: AF092506; AAC97365.1; -;

EMBL: AF092505; AAC98381.1; -
 DR MEROPS; S09.973; -
 DR MGD; MGI:94921; DPP6.
 DR InterPro; IPR002469; DPPV N term.
 DR InterPro; IPR001375; Peptidase_S9.
 DR Pfam; PF00930; DPPV N term; 1.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR Transmembrane; Glycoprotein; Signal-anchor.
 DR DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 35 55 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 56 804 EXTRACELLULAR (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 440 455 MISSING (IN REF. 1; AAC98381).
 FT CONFLICT 638 638 Q -> P (IN REF. 1; AAC97365).
 SQ SEQUENCE 804 AA; 91260 MW; 09CFCE7AD8A7168 CRC64;

Query Match 10.9%; Score 211; DB 1; Length 804;
 Best Local Similarity 19.7%; Pred. No. 3e-10;
 Matches 98; Conservative 65; Mismatches 159; Indels 176; Gaps 16;

QY 8 GVATFVLQEE-PRYSGYWMPKPAETTPSGGKILRLIYEENDESEVEIHHV---TSPMLE 63
 DB 226 GLSDWLYEEELIKSHIAHWMSPDG-----TRLAATINDSRVLMELPHYTSVVP 276
 QY 64 TRADSPRYKKTGANPKVTF-----KXSEIMI----- 91
 DB 277 T--VKRYHPKASSENSISLHVIGLNGPTHLEMPDDPRKREYIITVWKATSTKVA 334
 QY 92 -----DA-----EGR----- 96
 DB 335 VFWLNRQNVSLITLQDATTCVCKKHEDSEAMIRQNEBPFSQGRKFFVRAIPQG 394
 QY 97 -----IIIVDERLVYFEETKDSPLEHHLY 121
 DB 395 GRGKFYHITVSSSQPNSSNDNIQISGWDVTKILSYDEKRMKIFLSTEDLPRRHLY 454
 QY 122 VVSYNVGEVTRITLDGYSHSCISCHOPFISKYNSQKPHCVSLYKISSPEDPTCC-- 179
 DB 455 SANTVDDPNROCL-----SCDVENCTVVSASFHNMDD--FLLKCEGP-GVPTVTV 503
 QY 180 -----KTKFPAWITLDSAGPLPDYTPPEIFSPSESTGFTLYGMLYKPPHLDQ 226
 DB 504 HNTTDKRMEDLEANEVEQKAIINDRQMPKIEVKIEV-----EDYSLPMQILKPAFTFD 557
 QY 227 GKRYPTVLFTYGGPQVLVNRKGYKFFRLNTLASLGYVVVITDNGSCHRLKEGAF 286
 DB 558 TAHYPLLVVDDGTPGQSQVTEREPEVT--WETVIVSHGAVVVCDDRGSGGQCTKLQEV 615
 QY 287 KYMGQIISIDDQVEGLOYLARSYDFITLDRVGTHGWSYGGYSLMALMQRSD----IFRV 342
 DB 616 RRRIGLEBKQGEAVRMTL-KQYIDKTRAVVFGDYGISTYITLPAKGENQGGTFTFC 674
 QY 343 AIAGAPVTLMIIFYDGYT 360
 DB 675 GSALSPITDFKLKLYASAFS 692

RESULT 14

DPPE HUMAN STANDARD; PRT; 865 AA.
 AC P42658;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
 DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
 DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
 OS DPP6.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
 RC TISSUE=Hippocampus;
 RX MEDLINE=93372805; Pubmed=8103397;
 RA Yokotani N., Doi K., Wenthold R.J., Wada K.;
 RT "Non-conservation of a catalytic residue in a dipeptidyl
 aminopeptidase IV-related protein encoded by a gene on human
 chromosome 7.";
 RL Hum. Mol. Genet. 2:1037-1039(1993).
 CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
 CC ACTIVITY. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=DPPX-L;
 CC IsoId=P42658-1; Sequence=Displayed;
 CC Name=DPPX-S;
 CC IsoId=P42658-2; Sequence=VSP 005365;
 CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
 CC -!- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M96859; AAA35760.1; -
 DR EMBL; M96860; AAA35761.1; -
 DR PIR; I54331; I54331.
 DR MEROPS; S09.973; -
 DR GeneW; HANC:3010; DPP6.
 DR MIM; 126141; -
 DR GO; GO:0008239; F.dipeptidyl-peptidase activity; TAS.
 DR InterPro; IPR002469; DPPV N term.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR Pfam; PF00326; DPPV N term; 1.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
 DR DOMAIN 1 95 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 96 116 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 117 865 EXTRACELLULAR (POTENTIAL).
 FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 813 813 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1 81 NASTLYQRTFGKINTSRSPAPPEASHLLGGQEPEDGGAGA
 KPLGPAQAAPARBERGGGAGGAGRPFPYQGSDDDEED
 -> MTTAKEPSASGKSVQOOEOE (in isoform
 DPPX-S).
 FT FT /FTId=VSP 005365.
 SQ SEQUENCE 865 AA; 97588 MW; 14B1AB0E0024464B CRC64;

Query Match 10.5%; Score 203; DB 1; Length 865;
 Best Local Similarity 19.6%; Pred. No. 1.6e-09;
 Matches 98; Conservative 65; Mismatches 157; Indels 180; Gaps 17;

```

QY 8 GVATFVLOEE-FDRSYGWMCPKAEFTTSPGKILITLYEENDESEVEIHH---TSPMLE 63
DB 287 GLSDMLYEEELIKHTIAHMSPDG-----TRIAVAINDSVPMELPTTGTSTYR 337
QY 64 TRRADSFYRPKTGTANPKVTF-----KMSIIMI----- 91
DB 338 T--VKPYHYPKAGSENPSISLHVIGANGPTHDLBMMPPDDPRMREYITVMKMASTSKVA 395
QY 92 -----Da-----EGR----- 96
DB 396 VTMLNRAQNVSLITLCDATGTCVKRHEDESEAMLRHONREEPVSKORKEFTRAIPDG 455
QY 97 -----IIVDEVRLVYFEGTQDSCPLEHLY 121
DB 456 GNGKGYHITVSSSQPNSSNDNIQSTISGDMVTKLADENGNKITYLSTEDLRRRLY 515
QY 122 VSYVNPGEVTRLTDRGISH---SCCISQCHDFPISKYSNOKNPHCVSLYLSBP----- 173
DB 516 SAN-----TGNFNRQCLSCDLVENCITYFSASFHSDM--FELLKCEGPGVPMV 562
QY 174 -EDDPCTKTEF-----WATLDSAGPLDPYPRPIFSFESTGTLYGMLYKPHDL 224
DB 563 TVHNTTDDKKMFDELTHNEHVKKAINDROMPKVEYDII-----DYNLMQILKPRATF 616
QY 225 QPGKYPYVLFYGGPQVQVLYNNRPKGVKFRMLNTLASLGYVVVINDRSGCHRLKEEG 284
DB 617 TDTTYPLLLVVDGTPGSSQSAVEKEF--VSWETVWVSSHGAIVVVCDDRGSGFGOTKLH 674
QY 285 AKYMMGOIEIDDOVEGLOYLASRDYDILDRVGHGWSYGYSLMALMORSD----IF 340
DB 675 EVRRRLGLEEKQDQMEAVRMTL-KEQYIDRTKVAVFGKDYGYSLTYLTPAKGENGQOTF 733
QY 341 RVALIAGAPVTLMIFYDTGYT 360
DB 734 TCGSALSPITDFKLYASAFS 753

RESULT 15
YDZF SCHPO STANDARD; PRT; 853 AA.
ID YDZF SCHPO 013720;
AC Q9P7E9; 013720;
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Putative dipeptidyl aminopeptidase C14C4.15c (EC 3.4.14.-).
GN SPAC14C4.15C OR SPAP760.01C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxId=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Genies S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovich E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Wolchaert G., Aert R., Robben J., Grymoprez B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer B., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

```

```

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Sipakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature415:871-880(2002).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
CC vacuoles.
CC -1- SIMILARITY: Belongs to peptidase family 59B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: A1162631; CAB81084.1; -.
DR EMBL: Z98596; CAB1208.1; -.
DR GeneDB: SPombe; SPAC14C4.15c; -.
DR InterPro: IPR002469; DDPiv_N_term.
DR InterPro: IPR001375; Peptidase_S9.
DR Pfam: PF00930; DDPiv_N_term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
KM Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
KM Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 66
FT TRANSMEM 67 89
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 90 853
FT ACT_SITE 719 719
FT ACT_SITE 795 795
FT ACT_SITE 828 828
FT CARBOHYD 96 96
FT CARBOHYD 102 102
FT CARBOHYD 472 472
FT CARBOHYD 483 483
FT CARBOHYD 613 613
SQ SEQUENCE 853 AA; 98341 MW; 38450BA50F830486 CRC64;

Query Match 10.2%; Score 197.5; DB 1; Length 853;
Best Local Similarity 29.8%; Pred. No. 4.7e-09;
Matches 53; Conservative 25; Mismatches 59; Indels 41; Gaps 6;

QY 200 PPEIFSESTITGTYLGMLYKPHDLQPKKYPTVLFYGGPQVQVLYNNRPKGVKRYRLNT 259
DB 602 PSVFFKVIKYNITAIYQELRPENFNRKRYPTVFLYGAPOSALVGYKE---MDINE 657
QY 260 LASLGYVVVINDRSGCHRLKEGAFKXMGQIEIDPOVEG-----LQY 304
DB 658 LMASVNFVY-----KVDIVLSD--VSGQHLPFSDHELLIKSMWLL 698
QY 305 LASRYD-FRIDLRVGHGWSYGYSLMALMORSDIFRVAIAGAPVTLMIFYDTGYT 360
DB 699 LRSYVDTPTIDRHRVGHGWSFGGYLTL-KILENADPIRTGAVVAVPTDMRYDAYYS 755

Search completed: October 15, 2003, 17:11:34
Job time : 7.71145 secs

```



```

Db      61 MLETRADSFRRYKGTANPKVTFKXSEIMIDAEGRILIVDERRLVYEGKDSLEHNL 120
QY      121 YVSVYVNGEYRLTRDGRYSHSCCISQCHDFPISKYSNQNPHCVSLYKLSPEDDPTCK 180
Db      121 YVSVYVNGEYRLTRDGRYSHSCCISQCHDFPISKYSNQNPHCVSLYKLSPEDDPTCK 180
QY      181 TKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIOPGKKYPTVLFIYGGP 240
Db      181 TKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIOPGKKYPTVLFIYGGP 240
QY      241 QVOLVNNRFGKVKYFRNLNTLASLGYVVVVINDRSGCHRGKFEKAFKMGQIEIDDOVE 300
Db      241 QVOLVNNRFGKVKYFRNLNTLASLGYVVVVINDRSGCHRGKFEKAFKMGQIEIDDOVE 300
QY      301 GLQYLAHYDFIDIDRVGIGHMSYGYLSLMAIMQSRDIFRVALAGAPVTLMIFYDTGYT 360
Db      301 GLQYLAHYDFIDIDRVGIGHMSYGYLSLMAIMQSRDIFRVALAGAPVTLMIFYDTGYT 360

```

RESULT 2

```

Q8HBM5 ID Q8HBM5 PRELIMINARY; PRT; 882 AA.
AC Q8HBM5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Dipeptidyl peptidase 8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
RL Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL: AF221634; AAC29766.1; -.
DR MEROPS: S09.018; -.
DR InterPro: IPR002469; DPPIV_N_term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR000379; Ser esters_site.
DR Pfam: PF00930; DPPIV_N_term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
SQ SEQUENCE 882 AA; 101421 MW; AD80IC302DB4652B CRC64;

```

Query Match 95.0%; Score 1836.5; DB 4; Length 882;
 Best Local Similarity 67.5%; Pred. No. 8.4e-159;
 Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

```

QY      1 EEDARAGVATFVLQSEFDRYSGYVWCPKRAETTPSGKILRIIYEENDESEVEIIHVTSP 60
Db      244 EEDARAGVATFVLQSEFDRYSGYVWCPKRAETTPSGKILRIIYEENDESEVEIIHVTSP 303
QY      61 MLETRADSFRRYKGTANPKVTFKXSEIMIDAEGRILIVDERRLVYEGKDSLEHNL 120
Db      304 MLETRADSFRRYKGTANPKVTFKXSEIMIDAEGRILIVDERRLVYEGKDSLEHNL 363
QY      99 -----
Db      364 RAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERQRLIESVDSVTPLIY 423
QY      99 -----
Db      424 EETTDIWINIHDIHFVFPQSHHEBEIFIFASECKTGFRHLKYITSLIKESKYRRSSGGLP 483
QY      99 -----
Db      484 APSDFKCPKEIEIATSGEWEVLGRHGSNIQVDEVRRLVYFEGTQDSPLEHLLYVSYVN 543

```

RESULT 3

```

Q8IMG7 ID Q8IMG7 PRELIMINARY; PRT; 882 AA.
AC Q8IMG7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to dipeptidyl peptidase 8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC040203; AAH40203.1; -.
SQ SEQUENCE 882 AA; 101391 MW; 88C76AF5BCE707F9 CRC64;

```

Query Match 95.0%; Score 1836.5; DB 4; Length 882;
 Best Local Similarity 67.5%; Pred. No. 8.4e-159;
 Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

```

QY      1 EEDARAGVATFVLQSEFDRYSGYVWCPKRAETTPSGKILRIIYEENDESEVEIIHVTSP 60
Db      244 EEDARAGVATFVLQSEFDRYSGYVWCPKRAETTPSGKILRIIYEENDESEVEIIHVTSP 303
QY      61 MLETRADSFRRYKGTANPKVTFKXSEIMIDAEGRILIVDERRLVYEGKDSLEHNL 120
Db      304 MLETRADSFRRYKGTANPKVTFKXSEIMIDAEGRILIVDERRLVYEGKDSLEHNL 363
QY      99 -----
Db      364 RAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERQRLIESVDSVTPLIY 423
QY      99 -----
Db      424 EETTDIWINIHDIHFVFPQSHHEBEIFIFASECKTGFRHLKYITSLIKESKYRRSSGGLP 483
QY      99 -----
Db      484 APSDFKCPKEIEIATSGEWEVLGRHGSNIQVDEVRRLVYFEGTQDSPLEHLLYVSYVN 543
QY      128 RGEVTRLRDGRYSHSCCISQCHDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWAT 187
Db      544 RGEVTRLRDGRYSHSCCISQCHDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWAT 603
QY      604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIOPGKKYPTVLFIYGGPOVOLVNN 663
Db      188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIOPGKKYPTVLFIYGGPOVOLVNN 247
QY      604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIOPGKKYPTVLFIYGGPOVOLVNN 663
Db      248 RFGVKYFRNLNTLASLGYVVVVINDRSGCHRGKFEKAFKMGQIEIDDOVEGLQYLAS 307
QY      724 RYDFIDIDRVGIGHMSYGYLSLMAIMQSRDIFRVALAGAPVTLMIFYDTGYT 360
Db      724 RYDFIDIDRVGIGHMSYGYLSLMAIMQSRDIFRVALAGAPVTLMIFYDTGYT 360

```


OY 308 RYDFIDLRVGIHGSYGYSLSMALMQRSDIFRVAIAGAPVTLWIFDTGYT 360
 Db 724 RYDFIDLRVGIHGSYGYSLSMALMQRSDIFRVAIAGAPVTLWIFDTGYT 776

RESULT 4

ID Q9D4G6 PRELIMINARY; PRT; 892 AA.

AC Q9D4G6 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Adult male testis cDNA, RIKEN full-length enriched library,
 DE clone:493434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kamezaki T., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numasaki R., Ono M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai T.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
 RA Sojabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamuro T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Yamamatsu M., Hayashizaki Y.,
 RA Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Testis;
 RC MEDLINE=22354683; PubMed=12466851;
 RX The PANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN PANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690 (2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.,
 RT "High-efficiency full-length cDNA cloning."
 RL Meth. Enzymol. 303:19-44 (1999).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes."
 RL Genome Res. 10:1617-1630 (2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sakaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kitajima T., Tashiro T., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,

RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa Y., Tanaka T., Matsura S., Kawai J.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multichannel sequencer."
 RL Genome Res. 10:1757-1771 (2000).
 DR EMBL; AK016546; BAB30295.2;
 SO SEQUENCE 892 AA; 102284 MW; PDE0DBEDCA4CA346 CRC64;

Query Match 90.9%; Score 1756.5; DB 11; Length 892;
 Best local similarity 64.4%; Pred. No. 1.7e-151;
 Matches 343; Conservative 7; Mismatches 10; Indels 173; Gaps 1;

OY 1 EEDASAGATATVLDPEFPRVSGYVWCPKAEFTPSGCKTIRLYEENDESEVEIHHVSP 60
 Db 254 EEDPRAGATATVLDPEFPRVSGYVWCPKAEFTPSGCKTIRLYEENDESEVEIHHVSP 313
 OY 61 MLETRRADSPFRPKGTANPKVTFPMSEIMDAEGR11----- 98
 Db 314 MLETRRADSPFRPKGTANPKVTFPMSEIVDAAGI1DVKELVOPPELFEVGYIA 373
 OY 99 ----- 98
 Db 374 RAGWTPGKHAMSLIDRSQTHLQVLISPELFEVEDDAMDRORLIESVPSVPLI1Y 433
 OY 99 ----- 98
 Db 434 EETTDWINIHDFHVFPOTHEDLEFIFASCKTGPHLYKITSILKSKYKRSQGLP 493
 OY 99 -----VDEVRLLYFEGYKDSPLBHLVYVSYN 127
 Db 494 APSDFKCP1KEEITITSGEWEVLGHRGNIWDEARKVYFPGTDSPLBHLVYVSYN 553
 OY 128 PGEVTRLDGRGSHSCISQHCDFISKYSNOXNCHCVLSYLSPEDDPCKTEKFPAT 187
 Db 554 PGEVTRLDGRGSHSCISQHCDFISKYSNOXNCHCVLSYLSPEDDPCKTEKFPAT 613
 OY 188 IIDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLPQKKYPTVLFYGGPOVQLVNN 247
 Db 614 IIDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLPQKKYPTVLFYGGPOVQLVNN 673
 OY 248 REFQYKFRPLNTLASLGYVWVYIDNRGSHRDLKREGAFKYMGOIEIDDOVEGIQLAS 307
 Db 674 REFQYKFRPLNTLASLGYVWVYIDNRGSHRDLKREGAFKYMGOIEIDDOVEGIQLAS 733
 OY 308 RYDFIDLRVGIHGSYGYSLSMALMQRSDIFRVAIAGAPVTLWIFDTGYT 360
 Db 734 RYDFIDLRVGIHGSYGYSLSMALMQRSDIFRVAIAGAPVTLWIFDTGYT 786

RESULT 5

ID Q8NEMS PRELIMINARY; PRT; 831 AA.

AC Q8NEMS 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Homo sapiens (Human).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Strausberg R.,
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC030688; AAH30688.1; --
 SO SEQUENCE 831 AA; 95527 MW; 0B2A13A2FE70CB2 CRC64;
 Query Match 79.4%; Score 1534; DB 4; Length 831;
 Best local similarity 58.0%; Pred. No. 3.4e-131;

Matches 309; Conservative 0; Mismatches 0; Indels 224; Gaps 2;

```

QY -----BEDARSAGVATFVLQOEEDFDRYSQYWMCKAETTPSSGKILRIILYEENDESVEVILIHVTSIP 60
Db -----BEDARSAGVATFVLQOEEDFDRYSQYWMCKAETTPSSGKILRIILYEENDESVEVILIHVTSIP 303
QY -----61 MLETRRADSPRYEKTGTANPKVTFKMSIIMADAGRIL-----98
Db -----304 MLETRRADSPRYEKTGTANPKVTFKMSIIMADAGRILIVIDKELIQPELIFEGVEYIA 36
QY -----99-----98
Db -----364 RAGWTEBGKYAMSILLDRSQTRLQVILISELFIPEVDWMERORLIESVDSVTPLIYY 422
QY -----99-----98
Db -----424 EETTDIMINIHIDIFHWPOQSHSEIEIFIPASECKTGPRHLIKITSLIKESKYRASSGGLP 48
QY -----99-----127
Db -----484 ABSDECFPIKEBIAITSGSEWEVLGRHGSNIQVDEVRILYFEEGTOSPLEHHILVYVSYN 543
QY -----128 PGEVTRLTDRGYSHSCCISQCHDPFITSYNOQNPHCVSLYYKLSPEDDPCTCKTERMAT 187
Db -----544 PGEVTRLTDRGYSHSCCISQCHDPFITSYNOQNPHCVSLYYKLSPEDDPCTCKTERMAT 603
QY -----188 IIDSAGRLDYPPPELFSESTGTFLLYGMLYRPHDIQPEKRYPTVLFYIGSPQVOVLVNN 247
Db -----604 IIDSAGRLDYPPPELFSESTGTFLLYGMLYRPHDIQPEKRYPTVLFYIGSPQVOVLVNN 663
QY -----248 REKGYKFFRLNTLASIGYVVVVINDNGSGCHRGILFEGAFYKXMQGLEIDDQVBGLQYLIAS 307
Db -----664 REKGYKFFRLNTLASIGYVVVVINDNGSGCHRGILFEGAFYKXMQ-----706
QY -----308 RNDPFLDDBRVGIGHWSYGYLSLMLMQSNDIRPVVILAGAPVTLMIFYDTGYT 360
Db -----707 -----VALAGAPVTLMIFYDTGYT 725

```

```

QY 99 ----- 98
Db 146 RAGWTRDQKAYAMAFLEDRPQOMLOVLLPPALFIPSTENEBQRLASARAVPRVQPYVY 205
QY 99 ----- 98
Db 206 EEVTNWIMVHDIFFYPFQORERTSAFAPNARASAICTKSPPLKSGYDWESEPS 265
QY 99 ----- 98
Db 266 PGEDEFKCPKEBIALTSGEMEWLARHSGKIMWNEETKLVYFQGTOTPLEHLLYVVSYE 325
QY 127 NPGVETRLTDGYSGCCISQHCDFPISKYNOQKPHCVSLYKSSPEDDPTCKTEFWA 186
Db 326 AAGEIVRLTTPGFSHSCSMQNFDMFVSHSVSTPCVHYVLTSGPDDDLHKOFRWA 385
QY 187 TLIDSAGPLPDYTPPEIPEFSTGTFTLYGMLYKPHDLOPKKYPTVLFIYGPQVQVYN 246
Db 386 SMMEASCPDYPPEIPEFSTGTFTLYGMLYKPHDLOPKKYPTVLFIYGPQVQVYN 445
QY 247 NREKGVKPYRLNTLASIGYVVVYIDNRGSCRGKLFEGAFKYGQIEIDQVEGLQYLA 306
Db 446 NSFEGIKYRLNTLASIGYVVVYIDNRGSCRGKLFEGAFKYGQIEIDQVEGLQYLA 505
QY 307 SRYDFIDLDREVIGHWSYGYLSLALMQRSDIFRYALAGAPYTLFIYDGTGT 360
Db 506 EKVGFIIDLSRVAIHWSYGYLSLALMQRSDIFRYALAGAPYTLFIYDGTGT 559

```

RESULT 8

```

Q8WXD8 PRELIMINARY; PRT; 863 AA.
ID Q8WXD8
AC Q8WXD8;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DB Dipeptidyl peptidase 9.
GN DPE9.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Olsen C.; Wagtmann N.;
RT "Identification and characterization of a novel member of the
RT dipeptidyl peptidase IV-related family.";
RL Submitted (NCV-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA TISSUE=Skin;
RC Straussberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF452102; AL47179.1; -
DR EMBL; BC037948; AAH37948.1; -
DR MEROB; S09.019; -
DR InterPro; IPR002469; DPEV N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estr_site.
DR Pfam; PF00930; DPEV N term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
SQ SEQUENCE 863 AA; 98263 MW; 40FE0B7BE26CDED5 CRC64;

```

Query Match

```

Best Local Similarity 63.1%; Score 1220.5; DB 4; Length 863;
Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;

```

```

QY 2 EDARSAGVATFVLOEFPDRYSGYWCPCAKETTPSGG-KILRIYEENDESVEIITHVSP 60
Db 234 DPKSAGVATFVLOEFPDRYSGYWCPCAKETTPSGG-KILRIYEENDESVEIITHVSP 293
QY 61 MEETRADSPRYPKTGIANPKVTFKMSIIMIDAGRII----- 98

```

```

Db 294 ALEERKTDSPRYPKTGIANPKVTFKMSIIMIDAGRII----- 353
QY 99 ----- 98
Db 354 RAGWTRDQKAYAMAFLEDRPQOMLOVLLPPALFIPSTENEBQRLASARAVPRVQPYVY 413
QY 99 ----- 98
Db 414 EEVTNWIMVHDIFFYPFQORERTSAFAPNARASAICTKSPPLKSGYDWESEPS 473
QY 99 ----- 98
Db 474 PGEDEFKCPKEBIALTSGEMEWLARHSGKIMWNEETKLVYFQGTOTPLEHLLYVVSYE 533
QY 127 NPGVETRLTDGYSGCCISQHCDFPISKYNOQKPHCVSLYKSSPEDDPTCKTEFWA 186
Db 534 AAGEIVRLTTPGFSHSCSMQNFDMFVSHSVSTPCVHYVLTSGPDDDLHKOFRWA 593
QY 187 TLIDSAGPLPDYTPPEIPEFSTGTFTLYGMLYKPHDLOPKKYPTVLFIYGPQVQVYN 246
Db 594 SMMEASCPDYPPEIPEFSTGTFTLYGMLYKPHDLOPKKYPTVLFIYGPQVQVYN 653
QY 247 NREKGVKPYRLNTLASIGYVVVYIDNRGSCRGKLFEGAFKYGQIEIDQVEGLQYLA 306
Db 654 NSFEGIKYRLNTLASIGYVVVYIDNRGSCRGKLFEGAFKYGQIEIDQVEGLQYLA 713
QY 307 SRYDFIDLDREVIGHWSYGYLSLALMQRSDIFRYALAGAPYTLFIYDGTGT 360
Db 714 EKVGFIIDLSRVAIHWSYGYLSLALMQRSDIFRYALAGAPYTLFIYDGTGT 767

```

RESULT 9

```

Q8BWT9 PRELIMINARY; PRT; 862 AA.
ID Q8BWT9
AC Q8BWT9;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DB Dipeptidyl peptidase 9 homolog.
GN DPE9.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Liver;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK050021; BAC34034.1; -
SQ SEQUENCE 862 AA; 98001 MW; B1D566E824A834B8 CRC64;

```

Query Match

```

Best Local Similarity 62.3%; Score 1204.5; DB 11; Length 862;
Matches 234; Conservative 53; Mismatches 72; Indels 175; Gaps 2;

```

```

QY 2 EDARSAGVATFVLOEFPDRYSGYWCPCAKETTPSGG-KILRIYEENDESVEIITHVSP 60
Db 233 DPKSAGVATFVLOEFPDRYSGYWCPCAKETTPSGG-KILRIYEENDESVEIITHVSP 292
QY 61 MEETRADSPRYPKTGIANPKVTFKMSIIMIDAGRII----- 98
Db 293 ALEERKTDSPRYPKTGIANPKVTFKMSIIMIDAGRII----- 352
QY 99 ----- 98
Db 353 RAGWTRDQKAYAMAFLEDRPQOMLOVLLPPALFIPSTENEBQRLASARAVPRVQPYVY 412
QY 99 ----- 98

```

Db 413 EEVTNNWVNHDIHFHPPQAEAGOODFCFLANBCKTGFGHLKRVYVELTKDYDWTBPLS 472

QY 99 -----VDVBRLLVYFEGTKDSPLHLLVYVTV 126

Db 473 PTEDEFKCPKEEVALTSGEMEVLNRHSGSKIMVNEQTKLVYFQGTQDTLEHLLVYVSYE 532

QY 127 NPEVTVRLTRDGVSHSCCISQHCDFPISKYSNOKNHCVSLYKLSPEDDPCKTKEFMA 186

Db 533 SAGEIVRLTLTGFSHSCSMQSPFMYSHSVSTPFCVHVVYLDGSGCORGLHFGALKNMGQVELEDQVEGLQYVA 592

QY 187 TILDSAGBLPDYTPPELIFSESTGTGTLTGMLYKPHDLQPKKYPVTLFIYGGPOVOLVN 246

Db 593 SMMEANCPDYPPELIFHFTTRADVOLYGMVYKPHDLQPKKYPVTLFIYGGPOVOLVN 652

QY 247 NREKGVYFRNLNTLASIGYVVVYIDNRGSGHGLKPEGAFKYMGOEIDDOVEGLQYLA 306

Db 653 NSFKGKYLRLNTLASIGYVVVYIDNRGSGCORGLHFGALKNMGQVELEDQVEGLQYVA 712

QY 307 SRDYFDLDRVGIHGWMSGYGLSLMALMQRSDIFRVAIAGAPVTLFIYDTGYT 360

Db 713 EKGFIDLSRVALHGWMSGYGLSLMGLIHKQVFKVAILAGAPVTLVMAVDTGYT 766

RESULT 10

Q8BVG4 PRELIMINARY; PRT; 862 AA.

ID Q8BVG4

AC Q8BVG4; 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Dipeptidyl peptidase 9 homolog.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=olfactory brain;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."

RL Nature 420:563-573 (2002).

DR EMBL; AK078301; BAC37211.1; -

SQ SEQUENCE 862 AA; 98049 MW; B4FD3476B6F50030 CRC64;

Query Match 62.3%; Score 1204.5; DB 11; Length 862;
Best Local Similarity 43.8%; Pred. No. 4.6e-101;
Matches 234; Conservative 53; Mismatches 72; Indels 175; Gaps 2;

QY 2 EDARSAGVATFVLOEEDRYSYWCPCAKATTSSG-KILRIYENDSEVYIHWVSP 60

Db 233 DNPSSAGVATFVLOEEDRYSYWCPCAKATTSSG-KILRIYENDSEVYIHWVSP 292

QY 61 MLETRRADSPRYPKTGTPANPKVTFKMSIIMIDAEGRII----- 98

Db 293 ALBERKTDYRYRPTSGKNPKIKLAELQTDHOKTIVSSCEKLVQPPSSLPFKVYIA 352

QY 99 ----- 98

Db 353 RAGWTRDGYAMMFLYRPOORLQVLLPPLFIPAVESBACROAARAVPKVQPFVYI 412

QY 99 ----- 98

Db 413 EEVTNNWVNHDIHFHPPQAEAGOODFCFLANBCKTGFGHLKRVYVELTKDYDWTBPLS 472

QY 99 -----VDVBRLLVYFEGTKDSPLHLLVYVTV 126

Db 473 PTEDEFKCPKEEVALTSGEMEVLNRHSGSKIMVNEQTKLVYFQGTQDTLEHLLVYVSYE 532

QY 127 NPEVTVRLTRDGVSHSCCISQHCDFPISKYSNOKNHCVSLYKLSPEDDPCKTKEFMA 186

Db 533 SAGEIVRLTLTGFSHSCSMQSPFMYSHSVSTPFCVHVVYLDGSGCORGLHFGALKNMGQVELEDQVEGLQYVA 592

QY 187 TILDSAGBLPDYTPPELIFSESTGTGTLTGMLYKPHDLQPKKYPVTLFIYGGPOVOLVN 246

Db 593 SMMEANCPDYPPELIFHFTTRADVOLYGMVYKPHDLQPKKYPVTLFIYGGPOVOLVN 652

QY 247 NREKGVYFRNLNTLASIGYVVVYIDNRGSGHGLKPEGAFKYMGOEIDDOVEGLQYLA 306

Db 653 NSFKGKYLRLNTLASIGYVVVYIDNRGSGCORGLHFGALKNMGQVELEDQVEGLQYVA 712

QY 307 SRDYFDLDRVGIHGWMSGYGLSLMALMQRSDIFRVAIAGAPVTLFIYDTGYT 360

Db 713 EKGFIDLSRVALHGWMSGYGLSLMGLIHKQVFKVAILAGAPVTLVMAVDTGYT 766

RESULT 11

Q9BHM4 PRELIMINARY; PRT; 310 AA.

ID Q9BHM4

AC Q9BHM4; 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE Dipeptidyl peptidase 8 (Fragment).

GN DPP8.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=placenta;

RX MEDLINE=20467194; PubMed=11012666;

RA Abbot C.A., Yu D.M.T., Woolf E., Sutherland G.R., McCaughan G.W., Gorrell M.D.;

RT "Cloning, expression and chromosomal localization of a novel human dipeptidyl peptidase (DPP) IV homolog, DPP8."

RL Eur. J. Biochem. 267:6140-6150 (2000).

DR EMBL; AF221635; AAG29767.1; -

DR MEROPS; S09.018; -

DR InterPro; IPR001375; Peptidase_S9.

DR Pfam; PF00326; Peptidase_S9; 1.

FT NON TER 1

SQ SEQUENCE 310 AA; 35396 MW; BE87C34026D9C7AC CRC64;

Query Match 56.2%; Score 1086.5; DB 4; Length 310;
Best Local Similarity 80.6%; Pred. No. 6.7e-91;
Matches 204; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

QY 108 FEGTKDSPLHLLVYVSVNPGVTRTLDRGYSHSCCISQHCDFPISKYSNOKNHCVSL 167

Db 1 FEGTKDSPLHLLVYVSVNPGVTRTLDRGYSHSCCISQHCDFPISKYSNOKNHCVSL 60

QY 168 YKLSPEDDPCKTKEFWATILDSAGBLPDYTPPELIFSESTGTGTLTGMLYKPHDLQPG 227

Db 61 YKLSPEDDPCKTKEFWATILDSAGBLPDYTPPELIFSESTGTGTLTGMLYKPHDLQPG 120

QY 228 KKYPVTVFIYGGPOVOLVNNRFKVKYFRNLNTLASIGYVVVYIDNRSGCHGLKEGAFK 287

Db 121 KKYPVTVFIYGGPO----- 134

QY 288 YKMGQIEIDDOVEGLQYLAIRYFDLDRVGIHGWMSGYGLSLMALMQRSDIFRVAIAGA 347

Db 135 ---GQIEIDDOVEGLQYLAIRYFDLDRVGIHGWMSGYGLSLMALMQRSDIFRVAIAGA 191

QY 348 PVTLMIFDTGYT 360

Db 192 PVTLMIFDTGYT 204

RESULT 12

Q9BVR3 PRELIMINARY; PRT; 439 AA.

ID Q9BVR3

AC Q9BVR3; (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
 DE Hypothetical protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Strausberg R.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC000970; AAH00970.1; -.
 DR MEROPS; S09.019; -.
 DR Genew; HGNC:18648; DPP9.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR000379; Ser esters_site.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR Hypothetical protein.
 KW NON_TER
 FT NON_TER 1
 SQ SEQUENCE 439 AA; 49926 MW; A18BA9E12092BAF CRC64;
 Query Match 50.8%; Score 984; DB 4; Length 439;
 Best Local Similarity 64.8%; Pred. No. 2.5e-81;
 Matches 177; Conservative 40; Mismatches 56; Indels 0; Gaps 0;
 QY 88 EIMDAEGRIIVDEVRRLVFEETKDSPLFHHLYVSYVNGEYTRLTDRGSHSCCISQ 147
 DB 71 EVLARHGSKIVMEETLVYFGTKDPLFHHLYVSYEAAAGEIVRLTPGFSHSCMSQ 130
 QY 148 HCDEFISKYGNOKNPKCVSLYKLSPEDDPTCKTEPMATILDSAGPLPDYTPPEISFE 207
 DB 131 NFDMPFVSHSVSTPPCVHYKLSGPDDEPLHKQPRFMAEMEAASPPYVPELPHF 190
 QY 208 STTGFTLYGMLYKPHDLOPKKYPYVLFYGGPQVQLVNNRFKGVKXFRNLTLASIGYV 267
 DB 191 TRSDVRLYGMILYKPHALQPKKHTVLFYGGPQVQLVNNRFKGIKYLRLNTLASIGYV 250
 QY 268 VVIDNRGSGHKGKKEGAFYKMGQIEIDQVGLQYLAARYFIDLRGSHGMSGYG 327
 DB 251 VVIDGRSGCRGLFEGALKNQGVIEIDQVGLQYLAARYFIDLRGSHGMSGYG 310
 QY 328 LSLMALMQRSDIFRVALAGAPVTLMTFYDTGYT 360
 DB 311 LSLMGLIHKPQVFKVALAGAPVTVMAAYDTGYT 343
 RESULT 13
 ID 075273 PRELIMINARY; PRT; 508 AA.
 AC 075273;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
 DE R25984_1 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Lamerlin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
 Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
 Phan H., Velasco N., Do L., Regala W., Terry A., Gaines J.,
 Darganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
 Liu S., Altix C., Reutse T., Trachheim M., Amico-Keller G.,
 Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
 Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
 Kobayashi A., Olsen A.S., Carrano A.V.;
 RA "Sequence analysis of a 2.5 Mb region in 19p13.3."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC005594; AAC33801.1; -.
 DR MEROPS; S09.019; -.
 DR InterPro; IPR000379; Ser esters_site.
 FT NON_TER 1
 SQ SEQUENCE 508 AA; 57750 MW; 2FAD645BE2D2C89 CRC64;
 Query Match 48.8%; Score 944; DB 4; Length 508;
 Best Local Similarity 54.7%; Pred. No. 1.4e-77;
 Matches 185; Conservative 48; Mismatches 77; Indels 28; Gaps 5;
 QY 42 ILVENDSEVEIIVTSPMLTETRADSFRRY-----PKYGAN-PKVT-----FKMSE 88
 DB 75 VYVEEVTWVWVNHVIFFPFQSGEDDLCLRAEECTGICPLKYAVLAKSCGYDMS 134
 QY 89 IMIDAEGR-----IIVDEVRRLVFEETKDSPLFHHLYVSYVNGEYTRLTDRGSHS 142
 DB 135 PFSPEGQSOLTNALVMEETLVYFGTKDPLFHHLYVSYEAAAGEIVRLTPGFSH 194
 QY 143 CCIQGHDFISKYGNOKNPKCVSLYKLSPEDDPTCKTEPMATILDSAGPLPDYTPPE 202
 DB 155 CSMQNFDMFVSHSVSTPPCVHYKLSGPDDEPLHKQPRFMAEMEAASPPYVPELPHF 245
 QY 203 IFSFSTTGFTLYGMLYKPHDLOPKKYPYVLFYGGPQVQLVNNRFKGVKXFRNLTLAS 262
 DB 246 IFHFTSRDVRLYGMILYKPHALQPKKHTVLFYGGPQVQLVNNRFKGIKYLRLNTLAS 305
 QY 263 LGYVYVVDNRGSGHKGKKEGAFYKMGQIEIDQVGLQYLAARYFIDLRGSHGMSGYG 322
 DB 366 LGYVYVVDNRGSGCRGLFEGALKNQGVIEIDQVGLQYLAARYFIDLRGSHGMSGYG 365
 QY 323 SYGYVLSMALMQRSDIFRVALAGAPVTLMTFYDTGYT 360
 DB 366 SYGGLSLMGLIHKPQVFKVALAGAPVTVMAAYDTGYT 403
 RESULT 14
 ID 09HBM3 PRELIMINARY; PRT; 465 AA.
 AC 09HBM3;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE Dipeptidyl peptidase 8 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=20467194; PubMed=11012666;
 RA Abbott C.A., Yu D.M.T., Woollett E., Sutherland G.R., McCaughan G.W.,
 Gorell M.D.;
 RT "Cloning, expression and chromosomal localization of a novel human
 dipeptidyl peptidase (DPP) IV homolog, DPP8.";
 RL Eur. J. Biochem. 267:6140-6150 (2000).
 DR EMBL; AF221636; AAC29768.1; -.
 DR MEROPS; S09.018; -.
 KW NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 465 AA; 53197 MW; 22939EC0A4FE29CE CRC64;
 Query Match 44.5%; Score 860.5; DB 4; Length 465;
 Best Local Similarity 40.5%; Pred. No. 5.3e-70;
 Matches 186; Conservative 0; Mismatches 0; Indels 273; Gaps 2;
 QY 85 TGTAPKTYFMSEIMIDABGRIT----- 98
 DB 1 TGTAPKTYFMSEIMIDABGRITDVIDKLIQPELIFGVEYIARAGWTPEGKYAWSI 60
 QY 99 ----- 98
 DB 61 LLDQRQLQVILSPFLIPVDVMERQRLIESVPSVTPLIYBETTDIWINHIF 120

```

QY 99 ----- 98
Db 121 HVPQSHREIEFIASECKTGFRHLYKITSILKSKYKRSSGGLPAPSDKCPKEIEIA 180
QY 99 ----- -DEVPRIVYFEGTKDSPLIEHLVYVSVNPEVTRLDRGYSH 141
Db 181 ITSGEWEVLGRHGSNIQVDEVARLVYFEGTKDSPLIEHLVYVSVNPEVTRLDRGYSH 240
QY 142 SCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTEFWATILDSAGPLPDYTPP 201
Db 241 SCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTEFWATILDSAGPLPDYTPP 300
QY 202 EIFPSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQVQVQVNNRFGKVKYFRNLNTLA 261
Db 301 EIFPSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQVQVQVNNRFGKVKYFRNLNTLA 340
QY 262 SLGYVVVVIDNRGSGHRLKFEAGAFKYMGGQIEIDQVEGLQYLASRYDFIDLDRVGING 321
Db 341 ----- 340
QY 322 WSYGYSLSLMLMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
Db 341 ----- -VAIAGAPVTLWIFYDTGYT 359

```

RESULT 15

```

Q96NT8 . PRELIMINARY; PRT; 312 AA.
AC Q96NT8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ30094.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Nakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Magatsuna M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK054636; BAB70784.1; -.
DR InterPro; IPR001375; Peptidase_S9.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Hypothetical protein.
SQ SEQUENCE 312 AA; 35518 MW; ABE940AFCS877717 CRC64;

```

Query Match 41.3%; Score 799; DB 4; Length 312;

Best Local Similarity 66.2%; Pred. No. 1.3e-64;

Matches 143; Conservative 31; Mismatches 42; Indels 0; Gaps 0;

```

QY 145 ISOHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTEFWATILDSAGPLPDYTPPEIF 204
Db 1 MSQNDPMFVSHSYSTPCVAVVLYKLSGDDDDPLHKOPRFWSMMERASCPDDYVPEEIF 60
QY 205 SFESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQVQVQVNNRFGKVKYFRNLNTLASLG 264
Db 61 HHTHRSQVLYGMLYKPHDLQPKKYPVLYFYGGPQVQVQVNNRFGKVKYFRNLNTLASLG 120
QY 265 YVVVVVIDNRGSGHRLKFEAGAFKYMGGQIEIDQVEGLQYLASRYDFIDLDRVGINGWSY 324
Db 121 YAVVVVIDNRGSGHRLKFEAGAFKYMGGQIEIDQVEGLQYLAEKYGFTLSRVAIHGWSY 180
QY 325 GGYLSLMLMQRSDIFRVAIAGAPVTLWIFYDTGYT 360

```

```

Db 181 GGFSLIMGLIHKEQVFKVAIACAPVTVWMAVDTGYT 216
Search completed: October 15, 2003, 17:13:47
Job time : 25.0243 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:15:46 ; Search time 3930.37 Seconds
(without alignments)
3747.092 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEFPRD.....RVAIAGAPVTLMIFYDTGYT 360

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=x1h
-O=/cgn2.1/US10.spool/US10070464/runat.15102003.113553.24810/app.query.fasta_1.2652
-DB=GenBank1 -OPMT=fstestp -SUFFIX=rgs -MINMATCH=0.1 -LOOPCTL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi -LIST=45
-OCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US10070464.@CGN 1.1.1361.@runat.15102003.113553.24810 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenBank1.*
1: gb_da.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pri.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_scs.*
28: em_un.*

29: em_vl.*
30: em_htg_hum.*
31: em_htg_in.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1833	100.0	1083	9	AF221637 Homo sapi
2	1880	97.3	4309	6	AX608737 Sequence
3	1836.5	95.0	2649	6	AX354795 Sequence
4	1836.5	95.0	2656	9	AY172659 Homo sapi
5	1836.5	95.0	2671	6	AX608725 Sequence
6	1836.5	95.0	3106	6	AX342633 Sequence
7	1836.5	95.0	3127	9	AF221634 Homo sapi
8	1836.5	95.0	3143	6	AX354793 Sequence
9	1836.5	95.0	4535	6	BC040203 Homo sapi
10	1816.5	94.0	4829	6	AX608735 Sequence
11	1756.5	90.9	4799	10	BC043124 Mus muscu
12	1534	79.4	3130	9	BC030688 Homo sapi
13	1528	79.0	2161	6	BD157001 Primer fo
14	1528	78.0	2510	6	AK057826 Homo sapi
15	1522	78.7	2510	6	AX338497 Sequence
16	1522	78.7	4685	6	AX608745 Sequence
17	1521.5	78.7	4676	6	AX608743 Sequence
18	1254.5	64.9	2668	6	AX405771 Sequence
19	1254.5	64.9	2842	6	AX405770 Sequence
20	1230	63.6	2778	9	AK000290 Homo sapi
21	1229	63.6	4523	6	AX608731 Sequence
22	1220.5	63.1	2602	9	AY172660 Homo sapi
23	1220.5	63.1	2617	6	AX608727 Sequence
24	1220.5	63.1	3000	9	AX524928 Sequence
25	1220.5	63.1	3243	9	AF542510 Homo sapi
26	1220.5	63.1	3243	9	HS0605448 Homo sapi
27	1220.5	63.1	3716	6	AX480934 Sequence
28	1220.5	63.1	4219	6	AX608934 Sequence
29	1220.5	63.1	4232	9	AX608751 Sequence
30	1220.5	63.1	4295	9	AF452102 Homo sapi
31	1220.5	63.1	4302	6	AX608747 Sequence
32	1201	62.1	4180	6	AX608759 Sequence
33	1201	62.1	4263	6	AX608757 Sequence
34	1132.5	58.6	4076	6	AX608755 Sequence
35	1132.5	58.6	4159	6	AX608753 Sequence
36	1113	57.6	4037	6	AX608763 Sequence
37	1113	57.6	4120	6	AX608761 Sequence
38	1089	56.3	753	6	AX524935 Sequence
39	1088.5	56.2	1197	9	AF221635 Homo sapi
40	984	50.9	2546	9	BC000970 Homo sapi
41	975	50.4	2261	6	AX713363 Sequence
42	975	50.4	2261	6	AK054656 Homo sapi
43	860.5	44.5	1669	9	AF221636 Homo sapi
44	735	38.0	1080	6	AX524932 Sequence
45	718.5	37.2	3740	3	BT001499 Drosophila

ALIGNMENTS

RESULT 1

AF221637
LOCUS AF221637 1083 bp mRNA linear PRI 05-NOV-2000
DEFINITION Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, partial cds, alternatively spliced.
ACCESSION AF221637
VERSION AF221637.1 GI:11095193
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1083)
Abdott, C.A., Yu, D.M., Woollett, E., Sutherland, G.R., McCaughan, G.W. and Gorrell, M.D.
Cloning, expression and chromosomal localization of a novel human dipeptidyl peptidase (DPP) IV homolog, DPP8
Eur J Biochem. 267 (20), 6140-6150 (2000)
JOURNAL 20467194
MEDLINE 11012666
PubMed 11012666
REFERENCE 2 (bases 1 to 1083)
Abdott, C.A., Yu, D., McCaughan, G.W. and Gorrell, M.D.
Direct Submission
Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver
Centenary Institute of Cell Biology and Cancer Medicine,
Locked Bag No. 6, Newtown, Sydney, NSW 2042, Australia
FEATURES
source
1. .1083
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q22"
/clone="scPBMCD3-3-1"
/cell_type="PHA stimulated PBMC"
/gene="DPP8"
<1..>1083
/gene="DPP8"
/note="serine protease; member of prolyl oligopeptidase S9b family belonging to enzyme clan SC; additional splice variants of this mRNA are found in GenBank Accession Numbers AF221634, AF221635 and AF221636"
/codon_start=2
/product="dipeptidyl peptidase 8"
/protein_id="P46297.9.1"
/db_xref="GI:11095194"
/translation="BEDARAGVATFVLOEFPDYSGYWMCPRAEPTTSGKILRIILY EENDESFEIIVHTSPMLERADSEFRYPRTGTPANPYTFMSEIMIDAEGRITIDBEV RRLVREGTDSPLBHLVYVSVNPGEVRLIDRGVSHSCISQHCDFISKYGNL NPKCVSLPKLSSPEDDPCTCKEPMATILDSAGPLDIPYPEIISPSSTGTLTYGKL YKPHDLPGRKRTPLVLFYGGPOVLVNNPFKVKYFRRLTSLGIVYVVDNRSC HRKLKEGAFKRYKGOLEIDQVQGLVLSRYPDIDLDRVGIHWSYGGYLSLMALM QRSDFEVAIAGAPVTIMIFYDTGT"
BASE COUNT 312 a 213 c 247 g 311 t
ORIGIN
Alignment Scores:
Pred. No.: 4.24e-191 Length: 1083
Score: 1933.00 Matches: 360
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-070-464-7 (1-360) x AF221637 (1-1083)
QY 1 GluGIuAspAlaAGSerAlaGlyValAlaThrPheValIleuGInGluIuPheAspArg 20
Db 2 GAAGAAGATGCCAATAGCTGAGAGCTGCTTGTCTCCAGAAAGATTGGATAGA 61
QY 21 TyISerGIlyTYRTPTPCySProlYsAlaGluThrThrProSerGIlyGIlySIIleu 40
Db 62 TATCTGGCTATTGTGTGTGCCAAAGCTGAACAACCTCCAGTGGTGTAAATCTT 121

QY 41 ArgIIeLeuTYRGIuGIuAsnAspGIuSerGIuValGIuIIeIleHisValThrSerPro 60
Db 122 AGAATTCTATATGAAGAAATGATCAATCTGAGGTGGAAATATTATTCATGTTACATCCCT 181
QY 64 MetIleGIuThrArgArgAlaAspSerPheArgTYRProLYThrGIlyThrAlaAsnPro 80
Db 182 ATGTTGGAAACAGAGAGGCGAGATTCATTCGCTTATCTTAAACAGGTACACCAATCTCT 241
QY 81 LysValIThrPheLysMetSerGIuIleMetIIeAspAlaGIuGIlyArgIIeIleValAsp 100
Db 242 AAAGTCACTTTAAAGATGCAGAAATATGATGTAGTGTGGAAGGAAGATCAATAGTATGAT 301
QY 101 GIuValArgArgLeuValIlyPheGIuGIlyThrLysAspSerProIleuGIuHisIleu 120
Db 302 GAAGTCAGAGGCTGGTATATTTTGAAGGACCAAGACTCCCTTTAAGATCACTCG 361
QY 121 TYRValIValSerTYRValAsnProGIlyGIuValIThrArgLeuThrAspArgGIlySer 140
Db 362 TACGTATGTCAGTTCGTAATATCTGGAGAGGTGACAAAGCTGACCGTGGCTACTCA 421
QY 141 HisSerCYSPeSIIeserGIuHisCYAspPhePheIIeserIlySTYSerAsnGIuLys 160
Db 422 CATCTTGCTGCATCAGTCAGCAGCTGACTCTTTATAGTAAGTATGATTAACCAAGAG 481
QY 161 AsnProHisCYValSerLeuTYRlyLysLeuSerSerProGIuAspAspProThrCYsLys 180
Db 482 AATCCACACTGTGTGTCTCTTTACAGCTATCAAGCTCGAAGATACCACTTGGCAA 541
QY 181 ThrLysGIuPheTrpAlaThrIIeLeuAspSerAlaGIlyProIleuProAspTYRThrPro 200
Db 542 ACAGAGAAATTTGGGCGCACCATTTGGATTGACAGGTCCTCTCTCGTACTACTCTCT 601
QY 201 ProGIuIlePheSerPheGIuSerThrThrGIlyPheThrLeuTYRGIlyMetLeuTYRlys 220
Db 602 CCAGAAATTTCTCTTTTAAAGTACTACTGATGATTAATGATGAGAGCTCTACAG 661
QY 221 ProHisAspLeuGIuInProGIlyLysTYRProThrValIleuPheIIeTYRGIlyPro 240
Db 662 CCTCATGATCTACAGCTCGAAGAAATATTCATGCTGTGTTATATATGAGGTCT 721
QY 241 GlnValGIuLeuValAsnAsnArgPheLysGIlyValIlySTYRThrArgLeuAsnThrLeu 260
Db 722 CAGGTGCAGTGGTGAATATCGTTTAAAGAGTCAAGATTTCCGCTTGAATATCCCTA 781
QY 261 AlaSerLeuGIlyTYRValIValIleAspAsnArgLysCYSHISArgGIlyLeu 280
Db 782 GCCTCTCTAGTATATGTGTGATGATAGACAACAGGGGATCTGTCAACCGAGGCTCT 841
QY 281 LysPheGIuGIlyAlaPheLysTYRlySMetGIlyGlnIIeGIuIleAspAspGIuValGIu 300
Db 842 AATTTTGAAGGCCCTTTTAAATATTAATATGCTCAAAATTAATTAATTAATTAATTAAT 901
QY 301 GIlyLeuGIuIlyLeuAlaSerArgTYRAspPheIIeAspLeuAspArgValGIlyIleHis 320
Db 902 GAACCTCAATATATCTAGCTCTCGATATGATTAATTAATTAATTAATTAATTAATTAAT 961
QY 321 GIlyTrpSerTYRGIlyGIlyTYRLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePhe 340
Db 962 GGCCTGGCTTATGAGATACCTCTCTCTGATGACATTAATGACAGAGGTAGATATCTTC 1021
QY 341 ArgValAlaIleAlaGIlyAlaProValThrLeuTrpIlePheTYRAspThrGIlyTYRThr 360
Db 1022 AGGTTCTATTTGCTGGGCGCCAGTCACTCTGTGATCTTCTATGATACAGATACAG 1081
RESULT 2
AX608737 4309 bp DNA linear PAT 17-FEB-2003
LOCUS AX608737
DEFINITION Sequence 14 from Patent WO02211134.
ACCESSION AX608737
VERSION AX608737.1 GI:28404304
KEYWORDS
SOURCE Homo sapiens (human)

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Qi,S., Aktseanya,K.O., Riviere,P.J. and Junien,J.L.
TITLE	Novel serine protease genes related to dppiv
JOURNAL	Patent: WO 0231134-A 14 18-APR-2002;
FEATURES	Ferring BV (NL)
source	Location/Qualifiers 1..4309 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"
BASE COUNT	1304 a 790 c 907 g 1308 t .
ORIGIN	
Alignment Scores:	
Pred. No.:	7,46e-185
Score:	1880.00
Percent Similarity:	98.90%
Best Local Similarity:	98.90%
Query Match:	97.26%
DB:	Gaps: 0
US-10-070-464-7 (1-360) x AK608737 (1-4309)	
QY	1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
Db	943 GAAGAAGATGCGAATCAGCTGAGTGCCTACCTTGTCTTCCAAGAAATTTGATAGA 1002
QY	21 TyrSerGlyIYrTTPTPCyProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
Db	1003 TATTCGTGCTATTGGTGTGTGTCCAAAACCTGAACCACTCCAGTGGTGTAATCTT 1062
QY	41 ArgIleLeuYrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
Db	1063 AGAATTCATATGAAGAAATGATGAATCTAGAGTGGAATTTATTCATGTTACATCCCT 1122
QY	61 MetLeuGluThrArgArgAlaAspSerPheArgIYrPolySThrGlyThrAlaAsnPro 80
Db	1123 ATGTTGGAAACAGAGGAGGCGAGATTCATTCGGTATTCCTAAACACAGTACGACAAATCCT 1182
QY	81 LysValAlaThrPheLysMetSerGlnIleMetIleAspAlaGluGlyArgIleIleValAsp 100
Db	1183 AAGTCACCTTTAAGATGTCGAATTAATGATTGATGCTGAAGAGAAATCCAA-ATTGAT 1241
QY	101 GluValAlaArgArgLeuValTyrPheGlnGlyThrIlyAspSerProLeuGluHisIleLeu 120
Db	1242 GAAGTCAGAAAGCTGTATATTTTGAAGGACCAAGACCTCCCTTTAGACATCACTG 1301
QY	121 TyrValValSerTyrValAlaAsnProGlyGluValThrArgLeuThrAspArgIYrTyrSer 140
Db	1302 TACGTAGTCAGTTAGTAATCTTGGAAGGAGCAAGCTGACAGACCGGTGCTACCTCA 1361
QY	141 HisSerCysCysIleSerGlnHisCysAspPhePheIleSerIlyTyrSerAsnGlnLys 160
Db	1362 CATCTTGTGCTGATCAGTCAGACACTGTGACTCTTTAATAGTAATATAGTACACAGAG 1421
QY	161 AsnProHisCysValSerLeuTyrIlyLeuSerSerProGluAspAspProThrCysLys 180
Db	1422 AATCCACACTGTGTGTCCCTTTACAGCATCACTCCCTGAAGATGAGACCCAACTTGCAA 1481
QY	181 ThrIlyGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 200
Db	1482 ACGAAAGAAATTTGGGCCACCTTTTGGATTACAGAGGCTCTTCCCTGACTATATCTCT 1544
QY	201 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 220
Db	1542 CCAGAAATTTCTCTTTGAAGATGCTACTCGATTTAATGATGTATGGGATGCTCTCAAG 1601
QY	221 ProHisAspLeuGlnProGlyLysIlySerTyrProThrValIleuPheIleTyrGlyGly--P 240
Db	1602 CCTCATGATCTACAGCTGAGAAAGAAATATCTTACTAGGCTGTATATATAGAGTGCTC 1661

OY		240	rogluValGluLeuValAsnAsnArgPheylGlyValIlysTyrlPheArgLeuAsnThrL	260
Db		1662	CTCAGGGCGCAgTtGTGAATAAATCCGTTTAAGAAGACGACGATATTTCGGCTTGAAATACC	1721
OY		260	eulaSerLeuGlyTYrValValValValIleasPhasnArgIySerCySHisArgGLYL	280
Db		1722	TAGCCTCTCTAAGGTAAAGTGTGGTTGTAAGTAGAACAAAGAGGATCCTGTCCACGAGGC	1781
OY		280	eulySphegluglYAlaPheLyStrylsMetGLYcInll eglullleasPhsglnValG	300
Db		1782	TTAAATTTGAAAGGCCCTTTAAATATATAAATGGGTCAAATGAAATGAGCATCAGGTGG	1841
OY		300	IugLYLeuGlnITryLeuAlaSerArgTryspPheIleasPlaeuaspArgValGIyleH	320
Db		1842	AAGACATCCAAATATCTAGCTTCTCGAATATATTTCAATTGACTTAGATCGtGtGGCATTC	1901
OY		320	IsgLYTPSerTYrGlYGlYTyrLeuSerleuMetalaeMetGlnArgSerAspIleP	340
Db		1902	ACGGCTGTGCTCTAATGAGAGATACCTCTCCCTGATGGCATTAATGACAAGGTACATATCT	1961
OY		340	hearValAlaIleAlaIGLYAlAProValThrlleuTrpllePheTYraspThrGLYTYT	360
Db		1962	TCAGGGTGTGATTTGCTGGGGCCCCAGCTCACTGTGATCTTCTATGATATACAGATACA	2021
OY		360	hr 360	
Db		2022	CG 2023	
RESULT 3				
LOCUS	AX354795	2649 bp	DNA	linear PAT 06-FEB-2002
DEFINITION	Sequence 3 from Parent W00179473.			
ACCESSION	AX354795			
VERSION	AX354795.1	GI:18619528		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE				
AUTHORS	1 Meyers,R.A. and Williamson,M.			
TITLE	21953, a human prolyl oligopeptidase family member and uses thereof			
JOURNAL	Patent: WO 0179473-A 3 25-OCT-2001;			
	Millennium Pharmaceuticals, Inc. (US)			
FEATURES				
Source	Location/Qualifiers			
	1..2649			
	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
BASIS COUNT	803 a 514 c 586 g 746 t			
ORIGIN				
	1 36e-180 Length: 2649			
	Pred. No.: 1836.50 Matches: 360			
	Score: 67.54% Conservative: 0			
	Percent Similarity: 67.54% Mismatches: 0			
	Best Local Similarity: 95.01% Indels: 173			
	Query Match:			
	DB: 6 Gaps: 1			
US-10-070-464-7 (1-360) X AX354795 (1-2649)				
OY		1	GluGUAsPAalArxSeSerAlaGlyVAlaIaThrPheValLeuGlnGluGluPheAspArg	20
Db		730	GAGAGAAATGCCAAGATCACGCTGAGATCGCTTACCTTTGTTCTCCAGAGAAATTTGATACGA	789
OY		21	TyrSerGIyTYrTrpTYrCySProlYSAlaGluThrThrProSerGIyGLYLSlleu	40
Db		790	TATTCGTGCTATTTGGTGTGTCTCCAAAAGCTGAACAACACTCCACAGGTGTAAAAATCTCT	849
OY		41	ArgIleLeuTYrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro	60

Db AGATTCTATATGAGAAATGATGATCTGAGGTGAGAAATTATTCATGTTACATCCCT 909
 Qy 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProlyThrAlaAspPro 80
 Db 910 ATGTGGAAACAGAGGAGGAGATTCATCCGTTATCCCTAAACAGAGTACAGCAATCCCT 969
 Qy 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
 Db 970 AAGTCACTTTTAAGATGTCAGAAATTAATGATTGATGCTGAAAGAAAGATCATAGATGTC 1029
 Qy 98 ----- 98
 Db 1030 ATAGATTAAGAACTAATTCACCTTTTGAGATTCTATTGGAAGAGTTGAATATATGCC 1089
 Qy 98 ----- 98
 Db 1090 AGAGCTGATGAGCTCCGTGAGGAAATATGCTTGTCATCTACTAGATGCTCCAG 1149
 Qy 98 ----- 98
 Db 1150 ACTCGCTGCAATAGTGTGATCTGACCTGAATTATTATCCAGTAGAAGATGATGTT 1209
 Qy 98 ----- 98
 Db 1210 ATGAAAGGAGAGACTCATTTAGTCACTGCTGATTCGTGACGCCACTAATTATCTAT 1269
 Qy 98 ----- 98
 Db 1270 GAAGAAACACAGACATCTGATTAATATCCATGACATCTTTCATGTTTCCCAAGT 1329
 Qy 98 ----- 98
 Db 1330 CACGAAGAGAAATTAGATTATTTTGGCTCTGAATGCAAAACAGTTTCCCTCATTTA 1389
 Qy 98 ----- 98
 Db 1390 TACAAATTACATCTATTTTAAAGAAAGCAATATTAACGATCCAGTGGGCTGCT 1449
 Qy 98 ----- 98
 Db 1450 GCTCCAGTGAATTCAAGTGTCTTATCAAGAGAGATAGCAATTACAGTGGTGAATGG 1509
 Qy 99 ----- ValAspGluValArgArgLeuValTyr 107
 Db 1510 GAAGTCTTGCCGCGATGATCTAATATCCAGTTGATGAATCAGAAAGCGCTGATAT 1569
 Qy 108 PheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValValSerTyrValAsn 127
 Db 1570 TTGGAAGCACCAAGACTCCCTTAAAGACATCACCTGACGTACGTTACGTAAAT 1629
 Qy 128 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
 Db 1630 CCTGGAGAGGTGACAGGCTGACTGACCGGTGCTACCTCACATTTCTGCTGACATCAGTCA 1689
 Qy 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
 Db 1690 CACTGGACTTCTTTTAAGTAAGTATAGTAACCAAGAGATCCACACTGTGTGCTCCTT 1749
 Qy 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrPalaThr 187
 Db 1750 TACAGACTATCAAGTCTCTGAAGATGACCCCACTGCAAAACAAAGAAATTTGGGCGACC 1809
 Qy 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
 Db 1810 ATTTGATTCACAGGCTCTCTTCTGCTGACTATACCTCTCCAGAAATTTCTCTTTTGA 1869
 Qy 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
 Db 1870 AGTACTACTGATTTTACATTGATGAGTCTCTCAACAGCTCATGATCTACAGCTCGA 1929
 Qy 228 LysLysTyrProThrAlaLeuPheIleTyrGlyGlyProGlnAlaGlnLeuValAsn 247
 Db 1930 AAGAAATCTACTGCTGCTGCTATATATGATGAGTGTCTCAGGTGCGATTGGAAATAT 1989

Qy 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 Db 1990 CGGTTAAAGAGGATCAAGATATTCGCTGGAATACCTTAGCTCTAGGTTATGTGTT 2049
 Qy 268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyValaPheLys 287
 Db 2050 GTAGTATGACACACAGGGATCCTGTACCGAGGCTTAATTTGAAGGCCCTTTAA 2109
 Qy 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
 Db 2110 TATTAATGGGTCAAATAGAAATTCAGATCAGGTGAGAGACTCCAAATATCTAGCTTCT 2169
 Qy 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 327
 Db 2170 CGATATGATTTATGACTTATGATGCTGTGGCATCAGCGCTGCTGCTAGAGGATAC 2229
 Qy 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValaIleAlaGlyAla 347
 Db 2230 CTCTCCCTATGCGATTATATGACAGAGTCAAGATATCTTCAGGGTTCATTTGCTGGGCG 2289
 Qy 348 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
 Db 2290 CCAGTCACCTCTGATCTTCTATGATACAGGATACAG 2328
 RESULT 4
 AY172659 2656 bp mRNA linear PRI 08-JAN-2003
 LOCUS AY172659
 DEFINITION Homo sapiens dipeptidyl peptidase IV-related protein-1 (DPP1)
 ACCESSION mRNA, complete cds.
 VERSION AY172659
 KEYWORDS AY172659.1 GI:27549549
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
 TITLE Novel Serine Protease Genes Related to DPP1V
 JOURNAL Patent: US (WO 0231134) -A 18-APR-2002;
 2 (bases 1 to 2656)
 Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
 TITLE Direct Submission
 JOURNAL Submitted (04-NOV-2002) Ferring Research Institute, 3550 General
 Atomics Ct., San Diego, CA 92121, USA
 FEATURES
 source
 1..2656
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15q22.1-q22.2"
 1..2656
 /gene="DPP1"
 8..2656
 /gene="DPP1"
 /codon_start=1
 /product="dipeptidyl peptidase IV-related protein-1"
 /protein_id="AA017261.1"
 /db_xref="GI:27549550"
 /translation="MAAMETQLGVEIPETADCEENIESDRPKLEPFYVERYSWQ
 LKKLADTRKGVMAKAPHPFMVPRKNDPPGSDRIYVLAAMGSENNETLYSEI
 PKTRIAAVIMLSMKRPLDLPGATTDYCGYSEELIRRKRTITGYGASVYHOGSG
 TPLPQSGCITVYKKGPGQFTQQLPRNVLVETSCNIMDEPKLCPADPDWIAFIHNS
 DIMSNIVTRERRLTYVNELANMEEDARSAGATFVQLQEFDRYSGYWCPEKAIT
 PGGKTLRLYBENDESEVEIHHVTSPLMETRADSFYRKGTANPKYTFKMSLEIM
 DABGRLLDIVIDKELIOPFEILPEGVYILRAQMTPEGKYANSILDRSQRLQIVLS
 PELFIPVEDDVERORLIESVDSYPLIIEYETDININHIDIPVPOSHSEIEIF
 TRASECKGERHLRYITSLIKSKYRSGGGLPDSDFCPILKEEIAIRSGWEYVSG
 HSNINQVDEVRLLYFEGTKDSPLKHHLYVYSYVNPGEVTRITLDGYSKCCSCQSD
 FTRISKSNKPHCYSLKSSPEDDPTKTEFRNTILDSKGLDLDIYPPETFESES
 LTGFTLYGMLYKPHDLQPEKPTPLVFTYGGPQVQLVNNRFGVKYFRINTLASLGY

BASE COUNT 804 a 517 c 588 g 747 t
ORIGIN

Alignment Scores:

Pred. No.: 1,36e-180 Length: 2656
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: Gaps: 1

US-10-070-464-7 (1-360) x AV172659 (1-2656)

```

QY      1  GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluPheAspArg 20
DB      737  GAAGAAAGATGCCAATCAGCTGAGTGGCTACCTTGTCTCCAAAGAAATTGATAGA 796
QY      21  TyrSerGlyTyrTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleu 40
DB      797  TATTCTGCTATTGGTGGTGTCCAAAGCTGAAACAACTCCCACTGGTGGTAAATCTT 856
QY      41  ArgIleuTyrGlnGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
DB      857  AGAATCTTAATAGAAAGAAATGATGAACTCAGGTGGAAATTATTCATGTTACCTCCCT 916
QY      61  MetLeuGluThrArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
DB      917  ATGTTGAAACCAAGAGGCGAGATTCAATCCGTTATCCTAAACAGGTACGCAAAATCCT 976
QY      81  LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
DB      977  AAAGTCTCTTTTAAGATGTCAGAAATTAATGATTGATCTGAAGAGATCATAGATGTC 1036
QY      98  ----- 98
DB      1037  ATAGATAAGAACTAAATTCAACTTTGAGATCTAATTGAAAGATTGATATATTGCC 1096
QY      98  ----- 98
DB      1097  AGAGCTGAGTGAAGCTCCTGAGGAAATATGCTGTCCTACTACTAGATCGCTCCAG 1156
QY      98  ----- 98
DB      1157  ACTGCCTGCGAGATAGTGTGATCTCACTGAATTATTATCCAGTAGAAGATGATGTT 1216
QY      98  ----- 98
DB      1217  ATGGAAGAGCAGAGACTCATTTAGTCAAGTCCGATTCGTGAGCGCACATAATTATCTAT 1276
QY      98  ----- 98
DB      1277  GAAGAAACACAGACATCTGGATTAATATCCATGACATCTTTCAATGTTTTTCCCAAGT 1336
QY      98  ----- 98
DB      1337  CACGAGAGAGAAATTGAGTTATTTTGGCTCTGAATGCACAAACAGGTTCCGTCATTTA 1396
QY      98  ----- 98
DB      1397  TACAAATAATCATCTATTTTAAAGAAAGCAAAATATAACGATCCAGTGTGGCTGCT 1456
QY      98  ----- 98
DB      1457  GCTCCAAAGTATTTCAAGTGTCTTCAAGAAGAGATAGCAAAATTACAGTGTGAATGG 1516
QY      99  -----ValaspGluValArgArgLeuValTyr 107
DB      1517  GAAGTCTTGCGCGGAGATGATCTAATATCCAGATTATGAAGTCAGAGGCTGTATAT 1576

```

```

QY      108  PheGluGlyThrLysAspSerProLeuGlnHisIleuTyrValIleSerTyrValAsn 127
DB      1587  TTGAGAGGCAACCAAGACTCCCTTTAGACATCACCTGACGATCAGTTACGTAAAT 1636
QY      128  ProGluValValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGln 147
DB      1637  CTGGAGAGGTGACCAAGGCTGACTGACCGTGGCTACTCAACATTTCTGTGATAGTCAG 1696
QY      148  HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
DB      1697  CACTGTACTCTTCTTTAATAGTAAGTATAGTACCAAGAAATCCACACTGTGTCTCCTT 1756
QY      168  TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 187
DB      1757  TACAGGCTATCAAGTCTCTGAAGATGACCCCAACTGCAAAACAAAGAAATTTGGGCCACC 1816
QY      188  IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
DB      1817  ATTTTGATTCAGAGGTCCTCTTCTGACTATACCTCCAGAAATTTTCTTTTGAA 1876
QY      208  SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
DB      1877  AGTACTACTGATTTTACATTTGATGGAATGCTCTACAGCCTCATGATCTACAGCTCGA 1936
QY      228  LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
DB      1937  AAGAAATATCTCACTGCTGCTGCTCATATATGCTGCTCAGGTGCAAGTTGATAT 1996
QY      248  ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAsnSerLeuGlyTyrValVal 267
DB      1997  CGATTTAAGAGCAAGATTTCTCGCTTGAAATCCCTAGCCTCTCTAGGTATAGTGT 2056
QY      268  ValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluValAlaPheLys 287
DB      2057  GTAGTGATAGCAACAGGAGATCTCTGACCGAGGCTTAATTTGAAGGCGCTTTAA 2116
QY      288  TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
DB      2117  TATATAATGGGTCAAAATAGAAATGACGATCAGTGAAGAGCACTCAATATCTGCTTCT 2176
QY      308  ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyTyr 327
DB      2177  CGATATATTTTCATTTGACTTATGATCGTGGGACATCCACGCTGTGCTATGAGAGATAC 2236
QY      328  LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyVal 347
DB      2237  CTCCTCTGATGCGATTAATGACAGAGTCAAGATATCTTCAGGCTTGCTATTTGCTGGGCC 2296
QY      348  ProValThrLeuTyrPheTyrAspThrGlyTyrThr 360
DB      2297  CCACTCACTCGTGAATCTTCTATGATACAGAGATACAG 2335

```

RESULT 5
AX608725
LOCUS AX608725 2671 bp DNA linear PAT 17-FEB-2003
DEFINITION Sequence 2 from Patent WO0231134.
ACCESSION AX608725
VERSION AX608725.1 GI:28404238
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 OI.S., Aktinsanya, K.O., Riviere, P.J. and Junien, J.L.
Novel serine protease genes related to dppiv
Patent: WO 0231134-A 2 18-APR-2002;
Ferring BV (NL)
FEATURES
source 1..2671
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 805 a 524 c 594 g 748 t
ORIGIN

Alignment Scores:

Pred. No.: 1,37e-180 Length: 2671
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
Gaps: 1

US-10-070-464-7 (1-360) X AX608725 (1-2671)

QY 1 GIUGIUNSPALAAAGSERIAGLYVALAATHRPHVALLLENGINGIUNIPHEASPAR 20
Db 737 GAAGAAGATGCCAATCAGCTGAGTGGCTACCTTTGTCCTCAAGAGAATTGTGATGA 796
QY 21 TYSERGIYRTPCYSPROLYSALAGIUNTHRPHROSERGIYGLYSEU 40
Db 797 TATTCGCTATTGGTGTGTCCTCAAGCTCAAACTCCCACTGGTGTAAATCTT 856
QY 41 ARGILEUITYRGINGIUNASNPGLUSERGIUNVALGUILLEHISVALTHSERPRO 60
Db 857 AGAATTCATATGAAGAAATGATGAATCTGAGGTGGAATTTATTCATGATCCCT 916
QY 61 METLENGIUNTHRARGALAAASPSERPHENRGTYPOLYSTHNGIYTHRALASNP 80
Db 917 ATGTGGAAACAGAGGCGAGATTCATCCGTTATCTTAACACGGTACGAAATCT 976
QY 81 LYSVALTHRPHYSMETSERGIUNLEWETILEASPALAGIUNYARGILEU 98
Db 977 AAATCTCATTTAAGATGTGGAATATATGATTGCTGAAGAGGATCATGATGTC 1036
QY 98 ----- 98
Db 1037 ATAGATAAGAACTAATCACTTTGAGATCTATTGTAAGAGTTGAATATATGTC 1096
QY 98 ----- 98
Db 1097 AGAGCTGATGAGTCTCCTGAGGGAATAATGCTTGGTTCATCTAATGATGCTCCAG 1156
QY 98 ----- 98
Db 1157 ACTGCGCTGAGATAGTGTGATCTCACCTGAATATTATCCAGTAGAAGATGATGT 1216
QY 98 ----- 98
Db 1217 ATGGAAGGACAGAGCTCATTTGAGTCACTGCTGATCTGTGACGCCACTAATATCTAT 1276
QY 98 ----- 98
Db 1277 GAAGAAACACAGACATCTGATTAATATCCATGATCATCTTTTCCCAAGT 1336
QY 98 ----- 98
Db 1337 CACGAAGAGAAATGAGTTATTTTCTCTGANTCAAAACAGTTCCGTCATTTA 1396
QY 98 ----- 98
Db 1397 TACAAATATCATCTATTTTAAAGAAACAAATATTAACATCCAGTGGTGGCTGCT 1456
QY 98 ----- 98
Db 1457 GCTCAAGTATTTCAAGTGTCTCTCAAGAGGATGAGCAATTACAGTGGTGAATGG 1516
QY 99 ----- 98
Db 1517 GAAGTCTTGCCCGGATGATCTAATATCCAAAGTGTGATGAAGCTGGTATAT 1576
QY 108 PHEGIUNGLIYTHLYSASPSPROLENGIUNHISHSILEUITYRVALISERTYRVALASN 127
Db 1577 TTGTAAGGCAACCAAGACTCCCTTTAAGCATCTGCTGACGTGATGATGATTAAT 1636

QY 128 PROGIUNVALTHRARGLEUTHRAPARGIYTYRSEHISSERCYSIISERGIN 147
Db 1637 CCTGAGAGGATGACAGAGCTGACCTACCGTGGCTACTCATCTTCTGATCAGCAG 1696
QY 148 HISCYASPPHEPHEIIESELYTYRSEASNGIUNLYASNPROIHCYSVALSERLEU 167
Db 1697 CACTGTGACTCTTTAATAGTATAGTATAGTATACAGAAAGATCCAGCTGTGTCTT 1756
QY 168 TYRILEUSERSERPROGIUNASNPPTROTHCYSLYSHIRLYSGIUNPETHALATHR 187
Db 1757 TACAAAGCTATCAAGTCTCTGAAGATGACCCCACTTGCAAAACAAAGGAATTTGGGCCACC 1816
QY 188 ILEUASPSERIAGIYPROLEUPROASPTYRTHRPROGIUNILPEHSESPHEGLU 207
Db 1817 ATTTGGATTCAGAGGATCTCTCTGACTATACCTCCAGAAATTTCTCTTTGAA 1876
QY 208 SERTHRPHGLYPHETHRLEUTHYRGLYMETLEUITYRSPROHISASPLEUNPROGIY 227
Db 1877 AGTACTATCGATTTATCATTTGATAGGATGCTGTACAAAGCTCATGATCTACAGCTGGA 1936
QY 228 LYSLYSTYRPROTHRVALLEUPHEILETYRGLYGLYVPROGINVALGINLEUVALASNP 247
Db 1937 AAGAAATATCTCTGCTGTCTCATATATGTTGTTCTCAGGTGAGTGTGAATAT 1996
QY 248 ARGPHELYSGLYVALLYSTYRPHENRGLYASNPTRLEUNLASERLEUGIYTYRVAL 267
Db 1997 CGATTTAAAGAGCAAGTATTTCCGTTGAATACCTTAGCTCTCTAGTATGTGTT 2056
QY 268 VALVALILEASPAANARGIYSERCYHISARGIYLEUYSPEHGIUNVALPHELYS 287
Db 2057 GTAGTGTATGACCAACAGGGGATCCTGTACCGAAGGCTTAATTTGAAGGCGCTTTGAA 2116
QY 288 TYRILYMETGLYGINILEGUILLEASPSGLINVALGUNGILEUNTYRLEUASER 307
Db 2117 TATTAATATGGTCAATGAATGAATGAGATCAGTGAAGGACCTCAATATCTAGCTCT 2176
QY 308 ARGTYRASPHEIIEASPLEUASPARVALGILYIIEHISGLYTPSERTYRGLYGLYTYR 327
Db 2177 CGATATATTTCTATTTGATCTTATGATCTGTGGGATCCAGCGCTGTCTATGAGAGTAC 2236
QY 328 LEUSERLEUWETALAEUMETGINARSERASPILEPHENRVALALIEALGVALA 347
Db 2237 CTCTCCCTGATGGCATTAATGACAGAGTCAATATCTTCAAGGTTGCTATGCTGGGCGC 2296
QY 348 PROVALTHRLEUTHRPILEPHETRYSPHETHGLYTYRTHR 360
Db 2297 CCAGTCACTGTGTGATCTTCTATGATACAGATACAG 2335

RESULT 6
AX342633 3106 bp DNA linear PAT 12-JAN-2002
LOCUS AX342633
DEFINITION Sequence 30 from Patent WO0198468.
ACCESSION AX342633
VERSION AX342633.1 GI:18152030
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Yue,H., Bliotet,V.S., Gandhi,A.R., Lal,P., Au-Young,J.,
Tribouley,C.M., Deleogene,A.M., Baughn,M.R., Nguyen,D.B., Lee,E.A.,
Hafalla,A., Khan,F.A., Wallia,N.K., Yao,M.G., Lu,D.A., Patterson,C.,
Tang,Y.T., Walsh,R.T., Azimzai,Y., Ramkumar,J., Xu,Y. and Reddy,R.
Patent: WO 0198468-A 30 27-DEC-2001;
JOURNAL Inocyte Genomics, Inc. (US)
FEATURES
source
1..3106
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Inocyte ID No: 7160544CB1"

BASE COUNT 928 a 633 c 704 g 841 t

ORIGIN

Alignment Scores:

Pred. No.: 1,65e-180 Length: 3106
 Score: 1836.50 Matches: 360
 Percent Similarity: 67.548 Conservative: 0
 Best Local Similarity: 67.548 Mismatches: 0
 Query Match: 95.014 Indels: 173
 DB: 6 Gaps: 1

US-10-070-464-7 (1-360) x AX342633 (1-3106)

QY 1 GIUGIAspAlaArgSerIaGIyValAlaThrPheValLeuGlnGlnUpheAspArg 20
 DB 932 GAAGAAGATGCCAGATCAGCTGAGTGGCTACCTTGTCTCCAAAGAAATTTGATGATA 991
 QY 21 TYSerGIyTyrTrpTyrCysProLysAlaGlnThrThrProSerGIyGlyLysIleLeu 40
 DB 992 TATTCGGCTATTGGTGGTGTCCAAAGCTGAACACTCCAGTGGTGGTAAATTCCT 1051
 QY 41 ArgIleLeuTyrGlnGlnAspGlnSerGlnValGlnIleIleHisValThrSerPro 60
 DB 1052 AGAATTCATATGAAAGAAATATGATCATCTGAGGTGAATATTATCATGTATCATCCCT 1111
 QY 61 MetLeuGlnThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
 DB 1112 ATGTGGAAACAAGAGGAGGAGATTCATTCCTTATCTTAAACAGGTACAGCAAAATCCT 1171
 QY 81 LysValThrPheLysMetSerGlnIleMetIleAspAlaGlnGlyValArgIleIle 98
 DB 1172 AAGATCTATTTAAGATGTCAAGAAATATGATGATGCTGAGAGAGATCATGATGTC 1231
 QY 98 ----- 98
 DB 1232 ATAGATTAAGAACTAATTCACCTTTGAGATTCATTTGAAGAGTGAATATATTGCC 1291
 QY 98 ----- 98
 DB 1292 AGAGCTGATGATCTCCTGAGGAAATATGCTTGTGCATCTTACTAGATGCTCCAG 1351
 QY 98 ----- 98
 DB 1352 ACTCGCTACAGATAGTGTGATCTCACTCGAATTTATTCCTCAAGTAAGATGATGTT 1411
 QY 98 ----- 98
 DB 1412 ATGAAAGGAGAGACTCATTTAGTGCATGCTGATTCGTGAGCCACATAATATATCTAT 1471
 QY 98 ----- 98
 DB 1472 GAAGAAACAAGACATCTGATTAATATCATGACATCTTTCATGTTTTCCTCCAAAGT 1531
 QY 98 ----- 98
 DB 1532 CACGAAGAGAAATTGAGTTTATTTTTCCTCGATGCAAAAAGGTTTCGTCATTTA 1591
 QY 98 ----- 98
 DB 1592 TACAAATATACATCTATTTTAAAGAAAGCAAAATATTAAGATCCAGTGGTGGCTCCT 1651
 QY 98 ----- 98
 DB 1652 GCTCCAGTGAATTCATGATGCTCTCATCAAGAGAGATGCAATTACAGTGTGAATGG 1711
 QY 99 ----- ValAspGlnValArgArgLeuValTyr 107
 DB 1712 GAAGTCTTGGCCGAGATGATCTATATCCAAAGTTGAGAACTCAAGAGCTGGATAT 1771
 QY 108 PheGlnGlyThrLysAspSerProLeuGlnHisLysLeuTyrValValSerTyrValAsn 127
 DB 1772 TTGGAAGGACCAAGAACTCCCTTTAGAGCATCACTGTAGCTAGTACGTAAT 1831
 QY 128 ProGlnGlnValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGln 147

DB 1832 CCTGAGAGGTACCAAGCTGACTGACCTGTGCTACTCATCTTGTGCAATCATCTGAG 1891
 QY 148 HisCysAspPhePheIleSerLysTyrSerArgGlnLysAsnProHisCysValSerLeu 167
 DB 1892 CACTGTGACTTCTTTTATAGTATGATTAATACCAAGAAATCCACACTGTGTCTCTT 1951
 QY 168 TysLysLeuSerSerProGlnAspAspProHisCysLysThrLysGlnPheTrpAlaThr 187
 DB 1952 TCAAGACTATCAAGTCTCTGAAGATGACCCAACTTGGCAAAACAAGAAATTTGGGCACC 2011
 QY 188 IleLeuAspSerAlaGlyProLeuProAlaPyrTrpThrProGlnIlePheSerPheGln 207
 DB 2012 ATTTGGATTCAGCAGGTCTCTTCTGACTATATCTCCCAAAATTTTCTTTTGAA 2071
 QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
 DB 2072 AGTACTACATGATTTTACATTTGATGGAGATCTTACAACTCATATCTACAGCTGGA 2131
 QY 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 247
 DB 2132 AAGAAATATCTTACTGTCTGCTTCAATATGAGTGTCTCAGGTGCAATGATTAAT 2191
 QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAspSerLeuGlyTyrValVal 267
 DB 2192 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTAGGTATATGTGTT 2251
 QY 268 ValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGlnGlyAlaPheLys 287
 DB 2252 GTAGCATGACACAGGGAGATCCTGTACCCAGGGCTTAAATTTAAGCGCCTTTAA 2311
 QY 288 TysLysMetGlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnTyrLeuAlaSer 307
 DB 2312 TATTAATATGGTCAAAATGAAATGACATGACATGAGTGAAGAGACTCAATATCTTAC 2371
 QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyTyr 327
 DB 2372 CGATATGATTTATGACTTATGATCGTGGGCATCCAGGCTGTCTATGAGAGATAC 2431
 QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
 DB 2432 CTCTCCTATAGCATTAATGACAGAGTCAAGATATCTTCAAGGTCTTATGCTGGGCC 2491
 QY 348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
 DB 2492 CCAGTCACTGTGATCTTCTATGATACAGATACAGC 2530
 RESULT 7
 AF221634 3127 bp mRNA linear PRI 05-NOV-2000
 LOCUS
 DEFINITION Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, complete cds.
 ACCESSION AF221634
 VERSION AF221634.1 GI:11095187
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS 1 (bases 1 to 3127)
 Abbot, C.A., Yu, D.M., Woolliatt, E., Sutherland, G.R., McCaughan, G.W.
 TITLE Cloning, expression and chromosomal localization of a novel human
 dipeptidyl peptidase (DPP) IV homolog, DPP8
 JOURNAL Bur. J. Biochem. 267 (20), 6140-6150 (2000)
 MEDLINE 11012666
 PUBMED 20467194
 REFERENCE 2 (bases 1 to 3127)
 Abbot, C.A., Yu, D., McCaughan, G.W. and Gorrell, M.D.
 Epub. Submision
 Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver
 Centre, Centenary Institute of Cell Biology and Cancer Medicine,
 Locked Bag No. 6, Newtown, Sydney, NSW 2042, Australia
 FEATURES
 Location/Qualifiers

```

source      1. .3127
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /chromosome="15"
            /map="15q22"
            1. .3127
            /gene="DPP8"
            /gene="DPP8"
            /gene="DPP8"
            /note="serine protease; member of prolyl oligopeptidase S9b
            family belonging to enzyme clan SC; alternatively spliced
            forms of this mRNA are found in AF221635, AF221636 and
            AF221637"
            /codon_start=1
            /product="dipeptidyl peptidase 8"
            /protein_id="AAG29766.1"
            /db_xref="GI:11095188"
            /translation="MAAAMETBQLGVEIFETADCEENISQDRPKLEPPYVERYSNQ
            LKRLADTRKHGMMAKAPHDPMFYKNDPDGPHSDRIYLAISGENSENTLFTSEI
            PRTINAAVLMLSWKPLDLFOATLDYGMYSREELLRKRIGLVGLASYDHQSG
            TPLFOAGSGIYHVKGQPGQFTQQLRPNLVKSCNIMDPRLCPADPMIAFHSN
            DIWISNIVTREERRLYVHNELANMEDASAGVATFVLOEEDRSYGMWCPKAEIT
            PSQGLIRLILYENDESEVEIIVHTSPMLSTRADSFYKPGTANPKYTFPKSEIMI
            DAGRTIIVIDKELLQPELLEGGVEIVTARAGTPEKXAMGILLDRQTRLQIVLIS
            PELFIVEEDVHERKRLISVPSVPLIYETTDIMINIDIFRVFQSHBEIEF
            IPASBCKTGRHLYKITSILKESKYRSGGLPAPSDFCPIKEIETALTSGEHEVIGR
            HGSNIQDVNRRLVYEGTKDPSLEHLVVSVPNGVETRLDEIATLSGEHEVIGR
            FISKSNQKNPHCVSLYKLSPEDDPTCKEFMATILDSAGPLDYPPELFSPE
            TTGFTLGLYKHPDLPQKKYPTVLFIYGQVQLVNNRFGVKYERLNTLASIGYV
            VVVDNRGSCHRGLKREGAPKYMGQIEIIDDOVEGLCYLASRYDFIDLRVGHGMSY
            GGTLSLMALMQRSDFRVAIAGAPVTLMFTYDGYTERYMGHPDQEQYIGYSAMQ
            AEKPESEPRRLILHGFIDENVAHFATSLISFLVAKGPDQLIYPERHSIRVSES
            GHEYLHLHYIOMENLGSRIALXVI"

BASE COUNT      943 a      637 c      706 g      841 t
ORIGIN
Alignment Scores:
Pred. No.:      1,67e-180      Length:      3127
Score:          1836.50      Matches:      360
Percent Similarity: 67.54%      Conservative: 0
Best Local Similarity: 67.54%      Mismatches: 0
Query Match:     95.01%      Indels:      173
DB:              9      Gaps:      1

US-10-070-464-7 (1-360) x AF221634 (1-3127)
QY      1  GIUGLUSPALAARSERALAGLYVALALATHRPHENVALLLEUNGILUGLUPHEASPARQ 20
DB      943 GAAGAAGATGCCGATCAGCTGAGTGGCTACTTTGTTCTCAAGAAGATTGATAGA 1002
QY      21  TYRSEGLYTYRTTPCYSPOLYVALAGLINTHRPROSEGLYGLYLSILEU 40
DB      1003 TATTCGGCTATGTGGTGTCCAAAAGCTGAACAACCTCCAGTGTGTAATAATCTT 1062
QY      41  ARGILALEUTYGLUGLUSNASPGIUSERGLVALGILNLEHLSVALITHRSEPRQ 60
DB      1063 AGAATTCATATATAAAGAAATGATGATCTGAGGTGAATATTTCAITGTAATCCCT 1122
QY      61  METLEUGLINTHRARGTALAAASPERPHEARGTYRPROLYSTHRGLYTHRALAASPRO 80
DB      1123 ARGTTGAAAACAAAGAGGGCAGATTCATTCGTTATCCTAAACAGGTACAGCAAAATCCT 1182
QY      81  LYSVALITHRPHELYSETERGLIILEMELIASPALAGLUGLYARGILLE----- 98
DB      1183 AAAGTCACTTTAAAGATGTCAGAAATTAATGATGATGAGAGAAAGATCATAGATGTC 1242
QY      98  ----- 98
DB      1243 ATAGATAAGAACTAATTCACCTTTTGAGATTCATATTGGAAGAGTTGAATATATTGCC 1302
QY      98  ----- 98

```

```

DB      1303 AAGAGCTGATGAGACTTCCTGAGGGAAAATATGCTTGCTCATCTACTAGATGGCTCCAG 1362
QY      98  ----- 98
DB      1363 ACTGCGCTACAGATAGTGTGTGATCTCACTGAATATTATATCCAGTAGAAGATGATGTT 1422
QY      98  ----- 98
DB      1423 ATGGAAGGACAGACTCATTTGAGTCAGTCACTGATTCGTGACGCCAATAATTATCTAT 1482
QY      98  ----- 98
DB      1483 GAAAGAAACACAGACATCTGATTAATATTCATGACATCTTTCATGTGTTTCCCAAGT 1542
QY      98  ----- 98
DB      1543 CAGAAAGAAATTTAGTTTATTTTTCGCTGATGCAAAAACAGTTTCCGTCATTTA 1602
QY      98  ----- 98
DB      1603 TACAAATTTACATCTATTTTAAAGAAAGCAAAATATAACGATCCAGTGGCTGCCCT 1662
QY      98  ----- 98
DB      1663 GCTCCAGTGAATTTCAAGTGTCTATCAAGAGAGATAGCAATTACCAAGTGGTGAATGG 1722
QY      99  ----- 99
DB      1723 GAAGTTCCTGGCCGCGCATGATCTAATATCAAGATTGAAGTCAAGAGGCTGGATATAT 1782
QY      108 PHEGLUGLYTHRLYASPERPROLENGIUNHSILEUTYVALVASETYRVALASN 127
DB      1783 TTGGAAGGACAAAGACTCCCTTTAGAGCATCACTGATGATCACTGATTAACGTAAT 1842
QY      128 PROGLYGLVALITHRARGLEUTHIRASPARGLYTYRSEHISERCYSSILESEGIN 147
DB      1843 CCTGAGAGAGTGAACAAGCTGACGTACCGTGGTACTCACTCTTGCTGACATCAAGTCAG 1902
QY      148 HISYASPARPHEHELSEERYSTYRSERASNGILYASNPROIHSYVALSERLEU 167
DB      1903 CACTGGAATCTTTTATTAAGTATGATTAACCAAGAAATCCACACTGTGTGCTCTT 1962
QY      168 TYRLEULEUSERPROGLUSPASPPROTHRCYSLVETHRLYEGILUPHETPALATHR 187
DB      1963 TACAAAGCTATCAAGTCTGAAGATGACCCAACTGTGAAAACAAAGAAATTTTGGGCAACC 2022
QY      188 ILELEUASPERALAGLYPROLEUPROASPTYRTHRPROPROGLIUNLEPHESEPRHEGLU 207
DB      2023 ATTTGGATTTCAGCAGGTCTCTTCTGACTATACCTCCAGAAATTTTCTTTTGAA 2082
QY      208 SEETHRTHGLYPHETHRILEUTYRGLIYMETLEUTYRLYSPROHISASPLEUNGINPROGLY 227
DB      2083 AGTACTACACTGATTTTACATTTGATGGATGCTCTACMAACCTCATATCTACAGCCTGGA 2142
QY      228 LYSYSTYRPROTHRYALAEUPHELIETRYGILYGLYPROGLVALGILNEUVALAASNA 247
DB      2143 AAGAAATATCCACTGCTGCTGTTCATATATGAGTGTCCAGGTGAGTTGGATTAAT 2202
QY      248 ARGPHLYSGILVALIYSTYRPHEARGLLEUASNTHRLEUALASERLEUGLYTYRVAL 267
DB      2203 CGGTTTAAAGAGTCAAGATATTCCGCTTGAAATACCTTCAAGCTCTCTAAGTTATGTGGTT 2262
QY      268 VALVALIILEASPARNAARGLYSERCYSHLSARGLYLEULYRPHENGLUGLYALAPHELYS 287
DB      2263 GTAGATGATGACAAACAGGGGATCCGTGTCAACAGGCGTTTAAATTTGAAAGCGCTTTTAA 2322
QY      288 TYRLYSEMGILYGLNILEGLIULEASPARSGINVALGILUGLYLEUNGINTYRLEUALASER 307
DB      2323 TATTAATAGGTCGAATTAATAATTGACATACAGTGGAGAGACTCCAAATATCTAGCTTCT 2382
QY      308 ARGTYRASPPHEILEASPLEUNASPARGLYVALIILEHISGLYTYRSPERTYRGLYTYR 327
DB      2383 CGATATGATTTTCAATGACTTAGATCGTGTGGGATCCACAGGCTGTCTATGAGAGATAC 2442

```


QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
DB 2443 CTCTCCCTGATGGCATTATGACAGAGCTCAGATATCTTCAGGGTCTCTATTGCTGGGCGC 2502
QY 348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
DB 2503 CCAGTCACCTCTGGATCTTCTATGATACAGATACAGC 2541
RESULT 8
AX354793
LOCUS AX354793 3143 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 1 from Patent WO0179473.
ACCESSION AX354793
VERSION AX354793.1 GI:18619526
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Meyers, R.A. and Williamson, M.
AUTHORS 21953, a human prollyl oligopeptidase family member and uses thereof
TITLE Patent: WO 0179473-A 1 25-OCT-2001
JOURNAL Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
1. 3143
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
229..2877
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD22990.1"
/db_xref="GI:18619527"
/translation="MAAMETBOLGVEIFETADCEINISODPKLEPPYVERISNSQ
LKKLADTKYHGMYMAKAPHDFMFKNDPQSPHSDRIYLAISGENENTLFSBI
PKTINRAVLM,SWKPLDLFOATLDYGYSRBELRKRKIGVGIASYSYHQSG
TFPQAGSGIYHKDQGPQFTQOPLRPMLVTSCEINIMDEPLCPADPMIAFHISN
DIWISNIVTRESRLTYVNEELANMEEDASGAVATFVIOEEDRSYQWMCCKATY
PEGKRTILYENDESEVEIITHVSPMLSTRADSFRYPKGTANPKYTFKSRIMT
DABGRITIDIDKELIOPFEILBGEVYIARAGTPEGRKAMSLIDRSQTRIQIVLIS
PELFIPEDDVNERQRLIESVPSVPLTTEETDWINIHDFHVFQSHBEIEIF
IPASECKTPRHLKYITSLIKSKYKRSQGLPAPDFKCPKEELAIRSGEWEVGR
HGSNIQDVRRLVYEGTSDPLEHLLVVSYNVEYTRLDLRGSHSCCISOHCD
FISKSNQKNPCHVCSLYKLSPEDPTCKEPMATILDSAGPLDYPPELFSFES
TTGFTLYGMLYKPHDLPQKKYPTVLFIYGPQVQVNNRPFQVYERLNTALGLY
VYVIDRSGSCHRGKREKAFKMKQIETIDDOVBIQVLAASRYDPIIDLVRGIGHWSY
GGYLSLMALMGSDIRPRVALAAPTLMIFPYDTGYTERMGHPDQNEQYIAGSVAMQ
AKFPESEPRNLILHGFDENVHFAFTSILSLFVNAKRPYDLQIYPOBRHSIRVES
GERHELDLHYIOENLGRISIAALKVI"
BASE COUNT 943 a 644 c 712 g 844 t
ORIGIN
Alignment Scores:
Pred. No.: 1,68e-180 Length: 3143
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 6 Gaps: 1
US-10-070-464-7 (1-360) x AX354793 (1-3143)
QY 1 GlnGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
DB 958 GAAAGAGATGCCGATCAGCTGAGTGGCTACTTGTCTTCCAAAGAAATTGTATAGA 1017
QY 21 TyrSerGlyTyrTrpTyrCysProIysAlaGlnThrThrProSerGlyGlyIysIleLeu 40
DB 1018 TATTCTGCTATTGTGGTGTCCAAAGCTGAAACCACTCCAGTGGTAAATTCCT 1077
QY 41 ArgIleLeuTyrGlnGluLysAsnAspGluSerGluValGlnIleIleHisValThrSerPro 60

DB 1078 AGAATTCTATATGAAAGAAATATGATATGTAGGGGAAATTAATTCATATTCATCCCT 1137
QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProIysThrGlyThrAlaAsnPro 80
DB 1138 ATGTGGAAACAAAGAGGAGGACAAATTCATTCCTTATCTTAAACAGSTACAGCAATTCCT 1197
QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGlnGlyArgIleIle----- 98
DB 1198 AAGTCACCTTTTAAGTATGTCAGAAATTAATGATGATGATGAGGAAGATCATAGATGTC 1257
QY 98 ----- 98
DB 1258 ATAGATAGAGAACTAATTCACCTTTTGAGATCTTAATTGAAGAGTGAATATATTGCC 1317
QY 98 ----- 98
DB 1318 AGAGCTGATGACCTCTGAGGAAATATGCTTGGTCCATCTCACTACATGCTCCAG 1377
QY 88 ----- 98
DB 1378 ACTGCGCTGACAGTATGTTGATCTCACCCTGAATTATATCCAGTAAAGATGATGTT 1437
QY 98 ----- 98
DB 1438 ATGGAAGGACAGACTCATTTGATGATGATGCTGATTTCTGTAGCCCACTAATTATCTAT 1497
QY 98 ----- 98
DB 1498 GAAAGAAACAAACAGATCTGGATTAATATTCATGACATCTTATGTTTCCCAAGT 1557
QY 98 ----- 98
DB 1558 CAGGAAGAGAAATGATTTATTTTGGCTGTGAATGCAAAAAGGTTTCCGTCAATTTA 1617
QY 98 ----- 98
DB 1618 TACAAATTAATCATCTATTTTAAAGAAACAAATATATTAAGATCAGTGGGCTGCT 1677
QY 98 ----- 98
DB 1678 GCTCCAGGATTTCAAGTCTCTATCAAAAGAGAGATAGCAATTACCACTGGTGAATGG 1737
QY 99 -----ValAspGluValArgArgLeuValTyr 107
DB 1738 GAAGTCTTGGCGCGCATGATCTAATATCAAGTTGATGAACTCAGAAAGCTGGATAT 1797
QY 108 PheGlnGlyThrLysAspSerProLeuGlnHisIleLeuTyrValIysTyrValAsn 127
DB 1798 TTGGAAGGACACCAAGATCCCTTTAAGACATCACCTGTACGTACAGTTACGTAAAT 1857
QY 128 ProGlnGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
DB 1858 CCTGAGAGAGTACAAAGCTGACTGACCTGGCTACTCATCTTGTCTCATCAGTCAG 1917
QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
DB 1918 CACTGAGACTTCTTTTATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1977
QY 168 TyrIysLeuSerSerProGluAspAspProThrCysIysThrLysGlnIlePheIleThr 187
DB 1978 TACAAAGCTATCAAGTCTTCAAGATGACCAACTTGCAAAACAAAGAAATTTTGGGCAC 2037
QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGln 207
DB 2038 ATTTTGATTTACAGAGTCTCTCTCTGACTATCTCTCCAGAAATTTTCTTTTGA 2097
QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
DB 2098 AGTACTACGAGATTTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2157
QY 228 LysIysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 247

Db 2158 AAGAAATATCTACTGCTGTTTCATATATGTCCTCAGGTCAGTTGTAATAT 2217

QY 248 ArgPheLysGlyValIlystYrPheArgLeuAsnThleuAlaSerLeuGlyTyrValVal 267

Db 2218 CGGTTTAAAGAGACTCAAGTATTTCCGTTTAAATACCTTACCTCTTACGTTATGCGGT 2277

QY 268 ValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287

Db 2278 GTTGATATGACACACAGGAGATCTCTCACCAGGCGTTTAAATTTGAAGCGCTTTAA 2337

QY 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307

Db 2338 TATTAATATGGCTCAATATGAAATTTGACGATCAGGTGGAAGCATCCCAATATCTAGCTTCT 2397

QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyrGlyGlyTyr 327

Db 2398 CGATATGATTTCTATGATCTTATGATCGTGCGGATCCACGCGCTGCTTATGAGAGATAC 2457

QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347

Db 2458 CTCTCCCTGATGGCATTTAATGACAGGTCGATATCTTCAGGCTTGCTATGCTGGGGCC 2517

QY 348 ProValThrIeuThrIlePheTyrAspThrGlyTyrThr 360

Db 2518 CCAGTCACCTCTGTGATCTTCTATGATACAGATACACG 2556

RESULT 9

LOCUS BC040203 4535 bp mRNA linear PRI 03-DEC-2002

DEFINITION Homo sapiens, similar to dipeptidyl peptidase 8, clone MGC:42570

ACCESSION BC040203

VERSION BC040203.1 GI:26007916

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 4535)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (27-NOV-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: cgabs-rc@mail.nih.gov

Tissue Procurement: Miklos Palokovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Guaratene, P.H., Garcia, A.M., Lu, X., HuiLyn, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nananavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.lnl.gov>

Series: IRAC Plate: 71 Row: X Column: 23.

Location/Qualifiers

1..4535

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:42570 IMAGE:4824813"

/issue_type="Testis"

/clone_lib="NIH MGC_97"

/lab_host="DH10B"

CDS

/note="Vector: pBluescript"

1629..4277

/codon_start=1

/product="similar to dipeptidyl peptidase 8"

/protein_id="AAH40203.1"

/db_xref="GI:26007917"

/db_xref="LOCUSID:54878"

/translation="MAAMEPDLGVEIPEPADCEINSEODRPRKLEPPYVERYSMSO LKKLLADIKTHGYMAKAPHDPMYRKNDPPGPHSDRIYYLAMSSENENETIYSEI PXTINRAVALMISWKEPLDLFOATLDYGMYSSEELREKRRTIGVIGASVDYOSG TFLPQSGIYHVKGSGPOGFTQOPLRPMIVESCENIEMDPKICPADPMLAFIHGN DIWISNITVERERRLTYVENELANMEDRSAGVATFVIOEEDFRSGYWCCKAET PGSGKIRILYENENESVEIITHYSPMLETBADSFRYPKGTAPKTYFKKSEMTI DABGRIVYIDKEILOPRLIFRGVYITACGTPKCYKXAMSTILDRSOTRLOIVIS PELPIPEVDVNERQRILISVDSVTPILITETTDIWINIHIFVFPPOSHEEIEF IPASECKTGRFRLYKISTILKESKYRSGGLPAPSDPCPIKEIATISGEWELGR HSNANOVDVRLVYFEGTKDPLEHLIVVSVNGEVRLTDKRYSHSCISQHD FETSKSNQKNPCHVSLYKLSPEDEDPCTKEFMATILDSAGLPDPYTPPEIFSES TTGFTLYGMLYEPHLOPKKPTVLEFYGGPOVLNNRFKGVKFRLLNTLASLGY VVINDRSGCHRGKREGAFKYMGOIEIDDOVEGLQVYASRPDPLDLDVSGVMSO GGYLSMALMQSDIFRVALACAPITLMFYDGTERTERKMGHPDQVEQGYLGSVMO AKRPSPEPRLLHLHGFEDENVPATSTILSLFVAKGYPYDLOITPQERHSIRVES GRHYELHLHYIQENLIGSRIALKVI"

BASE COUNT 1313 a 918 c 946 g 1358 t

ORIGIN

Alignment Scores:

Pred. No.: 2,63e-180 Length: 4535

Score: 1836.50 Matches: 360

Percent Similarity: 67.54% Conservative: 0

Best Local Similarity: 67.54% Mismatches: 0

Query Match: 95.01% Indels: 173

DB: 9 Gaps: 1

US-10-070-464-7 (1-860) x BC040203 (1-4535)

QY 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluIupheAspArg 20

Db 2358 GAAGAAGATGCCAGATCACTGAGTCGCTACTCTTTGTTCCAAAGAATTTGATAGA 2417

QY 21 TyrSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40

Db 2418 TATTTGGCTATGTTGGTGTCCAAAGCTGAACAACCTCCAGTGGTGTAAATCTT 2477

QY 41 ArgIleLeuThrGluIupheAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 60

Db 2478 AGAATCTATATGAAGAAATATGATGATCTGAGCTGAAATATTTATGATACATCCCT 2537

QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80

Db 2538 ATGTTGGAACAAGAGGCGCAATTCATTCCTTATCTTAAACAGGTACAGCAATCCT 2597

QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleLeu 98

Db 2598 AAGTCACTTTTAAAGATGTCAGAAATATGATGATGTAAGAAAGATCATAGATGTC 2657

QY 98 ----- 98

Db 2658 ATGATTAAGAACTAATTTCACTTTTGAGATTCTATTGAAGAGTTGAATATATTGCC 2717

QY 98 ----- 98

Db 2718 AGAGCTGATGAGCTCTGAGGAAATATGTTGGTTCATCTCTACTAGATGCTCCAG 2777

QY 98 ----- 98

Db 2778 ACTGCGTACAGATAGTGTGATCTCAGCTGAATTTATTCACAGTAGAAGATGATGTT 2837

QY 98 ----- 98

Db 2838 ATGGAAGGACAGACTATGATGATGAGTCCGATTCTGTGACGCCACTAATTTATCTAT 2897

QY 98 ----- 98

Db 2898 GAAGAAACAACAGACATCTGGATMAATATCCATGACATCTTTCATGTTTTTCCCAAGT 2957
 QY 98 ----- 98
 Db 2958 CACGAAGAGAAATTGAGTTTATTTTGGCTCTGAATGCAGAAAGGTTTCCGTCTTAA 3017
 QY 98 ----- 98
 Db 3018 TACAAATTTACATCTATTTTAAAGAAAGCAATATAAAGATCCAGTGGTGGCTGCT 3077
 QY 98 ----- 98
 Db 3078 GCTCCAGTGAATTTCAAGTCTCTATCAAGAGAGATAGCAATTCACAGTGGTGAATG 3137
 QY 99 ----- ValAspGluValArgArgLeuValTyr 107
 Db 3138 GAAGTTCTTGGCCGGCATGGATCTAATCCAAAGTGAATGACAGAGGCTGGTATAT 3197
 QY 108 PheGluGlyThrLysAspSerProLeuGluHisLeuTyrValValSerTyrValAsn 127
 Db 3198 TTGGAAGGACCAAGAACTCCCTTAGAGCATCACTGTAGCTAGCTTACGTTACGTAAT 3257
 QY 128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
 Db 3258 CCTGGAGAGGTGCAAGAGCTGACGTGACCGTGGCTACTCACATTCCTGTGCATCAGTCA 3317
 QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
 Db 3318 CACTGCACTCTCTTTAAGTATAGTATAGTACCAAGAGATCCACATGCTGTGCTCT 3377
 QY 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
 Db 3378 TCAAGACTATCAAGTCTCGAAGATGACCCCACTTGCAAAACAAGGATTTTGGGCAC 3437
 QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 207
 Db 3438 ATTTGGATTCAGCAGGCTCTCTCTGACTATACCTCCAGAAATTTCTCTTTTGA 3497
 QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
 Db 3498 AGTACTACTGATTTTCACTTGATGAGATGCTTACAGCTCATATCTACAGCTCGA 3557
 QY 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 247
 Db 3558 AAGAAATATCTACTGCTGCTGCTCATATATGAGTGTCTCAGGTGCAAGTTGGTAAT 3617
 QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 Db 3618 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTAGGTTATGAGTT 3677
 QY 268 ValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
 Db 3678 GTAGTATAGCAACAAGGGGATCTGTCCAGAGGCTTAAATTTGAAGCGCTTTAAA 3737
 QY 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
 Db 3738 TATAAATGGTCAATTAAGAAATTCAGATCAGGTGAAGGAACTCCAATATCTAGCTTCT 3797
 QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyThrSerTyrGlyGlyTyr 327
 Db 3798 CGATATGATTTCACTTACCTTAGATCGTGGGATCCAGCGCTGGTCTCATGAGGATATC 3857
 QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
 Db 3858 CTCTCCCTGATGGCATTAATGACAGAGTCAAGATATCTTCAGGGTCTCATTTGCTGGGCC 3917
 QY 348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
 Db 3918 CCAATCACTCTGTGATCTTCTATGATACAGATACAGC 3956
 RESULT 10
 AX608735

LOCUS AX608735 4829 bp DNA linear PART 17-FEB-2003
 DEFINITION Sequence 12 from Patent WO0231134.
 ACCESSION AX608735
 VERSION AX608735.1 GI:28404303
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 Qi, S., Akinsanya, K.O., Riviere, P.J. and Junien, J.L.
 Novel serine protease genes related to dppiv
 Patent: WO 0231134-A 12 18-APR-2002;
 Ferring BV (NL)
 FEATURES
 source Location/Qualifiers
 1..4829
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 1466 a 886 c 1017 g 1460 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,41e-178 Length: 4829
 Score: 1816.50 Matches: 360
 Percent Similarity: 67.29% Conservative: 0
 Best Local Similarity: 67.29% Mismatches: 0
 Query Match: 93.97% Indels: 175
 DB: 6 Gaps: 1
 US-10-070-464-7 (1-360) x AX608735 (1-4829)
 QY 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluIlePheAspArg 20
 Db 943 GAGAGAGATGCCAGATCACTGAGTGGCTGCTACTTGTCTCCAGAGAAATTTGATAGA 1002
 QY 21 TyrSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
 Db 1003 TATTCGGCTATGTTGGTGGTGTCCAAAGCTGAAACAACCTCCAGTGGTGAATATCTT 1062
 QY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleHisValThrSerPro 60
 Db 1063 AGAATTCATATGAAAGAAATGATGATCTGAGTGGAAATTAATTCATGTTACATCCCT 1122
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAspPro 80
 Db 1123 ATGTTGGAACCAAGAGGGGCAATTCATTCGTTATCTTAAACAGGTACAGCAATCTCT 1182
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
 Db 1183 AAAGTCATTTTAAAGATGTCAAGAAATTAATGATGCTGAAGAGGATCATGATGTC 1242
 QY 98 ----- 98
 Db 1243 ATAGATAGAGAACTAATTCACCTTTTGAGATTCATTTGAGAGAGTGAATATATTGCC 1302
 QY 98 ----- 98
 Db 1303 AGAGCTGATGAGACTCTTGAAGGAAATATGCTGTGTCATCTTACATGATCGCTCCAG 1362
 QY 98 ----- 98
 Db 1363 ACTCGCTACAGATAGTGTGATCTTCACCTGAATTAATTAATCCAGTAGAAGATGATGTT 1422
 QY 98 ----- 98
 Db 1423 ATGAAAGGACAGACTCATTAAGTCAGTGGCTGATTCGTGAAGCCACTAATTAATTCAT 1482
 QY 98 ----- 98
 Db 1483 GAAGAAACAACAGACATCTGATTAATATCCATGACATCTTCATGTTTTCCTCCCAAGT 1542
 QY 98 ----- 98

```

Db      1543 CACGAAGAGAAATTGAGTTATTTTTGGCTCTGAATGCATAAAGAGTTTCCGTCATTTA 1602
QY      98 -----
Db      1603 TACAAATTATCATCTATTTTAAAGAAAGCAAAATATAAACGATCCAGTGGTGGCTGCCT 1662
QY      98 -----
Db      1663 GCTCCAGTGAATTCAGAGTCTCTATCAAGAGAGATAGCAATTAACAGTGGATAGG 1722
QY      99 -----
Db      1723 GAAGTTCTGGCCGGCATGATCTAATATCCAGTTGATGATCACTCAGAGGCTGGATAT 1782
QY      108 PheGluGlyThrIysAspSerProLeuGluHisIleuTyValIleSerTyValAsn 127
Db      1783 TTGGAAGGACCAAAAGCTCCCTTTAGAGCATCACCTGACCTAGTACGTTACGTAAT 1842
QY      128 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
Db      1843 CCTGGAGAGGTGACAAAGCTGACGACCGTGGCTACCATCTTGCTGCATCAGTCAAG 1902
QY      148 HisCysAspPhePheIleSerIleTySerAsnGlnIleAsnProHisCysValSerLeu 167
Db      1903 CACTGAGCTTCTTTTAAGTAACTAGTATGTAACCAAGAGATCCACACTGTGTGCTCCT 1962
QY      168 TyrIleLeuSerSerProGluAspAspProThrCysIleThrIleGluIlePheTyrPalaThr 187
Db      1963 TACAGCTATCAAGCTCCTGAAGATGACCCCACTTGCAAAACAAAGGATTTTGGGCCACC 2022
QY      188 IleLeuAspSerIleGlyProLeuProAspTyrThrProProGluIlePheSerPheGln 207
Db      2023 ATTTGGATTCACAGAGTCTCTCTCTGACTATACCTCTCAAAATTTCTCTTTGAA 2082
QY      208 SerThrThrGlyPheThrLeuTyrglyMetLeuTyrIleSerProHisAspLeuGlnProGly 227
Db      2083 AGTACTACTGAGATTTCATTGTATGGGATGCTCTTACAAGCTCATGATCTACAGCCTGGA 2142
QY      228 LysIleTyProThrValLeuPheIleTyrGlyGly--ProGlnValGlnLeuValAsn 247
Db      2143 AAGAAATATCTCTACTGCTGCTGCATATAATGAGTGCTCTCTCAGGCTGAGTGTGTAATA 2202
QY      247 snArpPheIleGlyValIleTyThrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValIle 267
Db      2203 ATGGGTTTAAAGAGTCAAGTATTTCCGCTTGATACCTTACGCTCTCTAGGTTATGTGG 2262
QY      267 alValValIleAspAsnArgGlySerCysHisArgGlyLeuIleGlyPheGluGlyValaPheIle 287
Db      2263 TTGTAGATAGACAAACAGGGGAGCTGTCACCGAGGGCTTAATTTGAAGGCGCCTTAA 2322
QY      287 YefTyrIleMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaS 307
Db      2323 AATATATAAAGGGTCAAAATGAAATGACGATGAGGAGAGGACTCCCAATATCTAGCTT 2382
QY      307 eArGlyTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPseTyrGlyGlyT 327
Db      2383 CTGCAATGATTTCACTTGACTTGAATCGTGTGGCAATCCACGGCTGTCTATGAGAGAT 2442
QY      327 YrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 347
Db      2443 ACCCTCCCTCGATGGATTAATGACAGGCTCAGATATCTTCAAGGTTGCTATATGCTGGGG 2502
QY      347 laProValThrLeuTyrIlePheTyrAspPheThrGlyTyrThr 360
Db      2503 CCCAGTCACTGTGTGATCTCTATGATACAGAGATACAG 2543

```

```

RESULT 11
BC043124      4799 bp      mRNA      linear      ROD 22-APR-2003
LOCUS      BC043124
DEFINITION      Mus musculus dipeptidyl peptidase 8, mRNA (cDNA clone MGC:58057
IMAGE:6410075), complete cds.
ACCESSION      BC043124

```

```

VERSION      BC043124.1 GI:27695449
KEYWORDS      MGC.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4799)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buco, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, U., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Frange, C., Rana, S.S., Loggellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
Mckernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kerteman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Scherchenko, Y.,
Boutard, G.G., Blakeley, R.W., Touchman, D.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Gutierrez, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E.,
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2 (bases 1 to 4799)
Strausberg, R.
Direct Submission
Submitted (09-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NHI-MGC Project URL: http://mgc.ncl.nih.gov
REMARK      MGC help desk
COMMENT      Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas U. Casavant.
Web site: http://genome.uiowa.edu
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldo, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,
Fisher, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
Scheetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J.,
Casavant, T., Soares, M.B.
Clome distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: Plate: Row: Column: 0.
Location/Qualifiers
1..4799
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/db_xref="MGC:58057 IMAGE:6410075"
/cisue_type="Brain, enriched mouse brain 12.5dp"
/clone_lib="NHI BMAP_F00"
/lab_host="DH10B"
/note="Vector: pYX-ASC"
1..4799
/gene="Dpp8"
/db_xref="LocusID:74388"
/db_xref="MGI:1921638"
288..2966
/codon_start=1
/product="Dpp8 protein"

```

/protein_id="AAH43124.1"
/db_xref="GI:27695450"
/db_xref="LOCUSID:74388"
/translation="MKIPSGRCNMAAAMETQIAGVIEFAECBEGNGESODRPLER
FYVERYSWSOLKLIADTRKYGYMAAPHDPMFMTGDPGMSDLYVYLAMGEN
RENTLFYSBEPKLTINRAVIMLSMKPLDLPDADYDMSSEELDERKICVGI
AAYDHPSSGTFLEFQAGSITHIKDGHGFTQPLRNVLVETSCPNRMKCPAD
PDWAFHISNDWISNLVREERITVYELNANMBDPRSGVATFVLOEFPDYG
YMCPOARSPSGKILRLIYEENDESEVLIHVTSPLMETRADSPRYPKGTANPK
VTFKSEIIVDPAAGIIVIDKEILOPELIEGVEYIARAGMTEGHGMSILDRS
OTLOLIVISPELFIPEVDAMDRORLIESVPSDVPLIYEETDWINIHIDHVF
POTHEDELEFIASBCKTGPRHLVTKITSLIKSKRSBGGLPASPDKCPKEKITI
TSGEMVTCRHOSNITWDBARKLVIFEGTKDSPLFHHVTSYANPGEVRLTDGYS
HSCCLSRHCDPEFISKYSNOKNPCVSLKSLSPEDPVHKTKEFWATLIDSGPLPDY
TPPELFESFESTTGFTLYGMLYKPHDLOPKKPYLVLFYGGQVQLVNNRFGYVFR
INTLASLGYVVVVDNRGSCHRGLKFEAGFKYKMOIHIIDQVEGLQYIASQYDIDL
DRVGIHWSYGGYSLMALMORSDFRVALIGA PVTLMIFDTGTERTYEMGHQDNO
CYLGSVAMQAEKRPSPRRLILHGFIDENHPAHTSILSLFLVRAGKPDLOIYPO
ERHSIRVPSGSEHRLHLYIQENLGSRIALAKVI"

BASE COUNT 1373 a 1015 c 1103 g 1308 t
ORIGIN

Alignment Scores:
Pred. No.: 5 81e-172 Length: 4799
Score: 1756.50 Matches: 343
Percent Similarity: 65.67% Conservative: 7
Best Local Similarity: 64.35% Mismatches: 10
Query Match: 90.87% Indels: 173
DB: 10 Gaps: 1

US-10-070-464-7 (1-360) x BC043124 (1-4799)

QY	1	GIJUAASPALAEISERIALGLVYALATHRPHEVALLEUGLUGLUPHEASPARG	20
DB	1047	GAAGAGGATCCAGATCAGCTGCTGTGCACTTTGTTCTTCAAGAAATTTACACGA	1106
QY	21	TyrSerGlyTyrTyrTyrCysProlysalagluThrProSerGlyGlyLysIleLeu	40
DB	1107	TACTGTGCTACTGTGTGTGTCCCAAGCGGAAAGACTCTTAGTGATGTAATTCCTT	1166
QY	41	ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro	60
DB	1167	AGAAATTCCTTATGAAGAAATATGATATGTGAGTGAGATTAATTCATGTCGCCCC	1226
QY	61	MetLeuGluThrArgArgAlaAspSerPheArgTyrProlysthrGlyThrAlaAsnPro	80
DB	1227	ATGTTGAAACACAGAGGCGACATTCCTTCGTTATCCCAAAACAGGCACAGCAACCA	1286
QY	81	LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle-----	98
DB	1287	AAGTCACCTTCAAGATGTCGAGATGTTGTGTATGTCGACGAGAGGATTTAGATGTC	1346
QY	98	-----	98
DB	1347	ATAGATAAAGAACTGGTTCAACCTTTCAGATTCCTGTTGAGGGAGTGAATATATTGCC	1406
QY	98	-----	98
DB	1407	AGAGCCGATGACTCCAGAGGAAACATGCGTGCATCTACTAGAACCGTTCACG	1466
QY	98	-----	98
DB	1467	ACTACCTGCAGATAGTTCTGATCTCCCTGAGTTATTCATCCAGTAGAAGATGATGCC	1526
QY	98	-----	98
DB	1527	ATGACACAGACAGACTTATAGAGTCAGTTCCTGACTCTGTACACACCATGATCATTTAT	1586
QY	98	-----	98
DB	1587	GAAGAAACACAGACATCTGATATAATATCCAGATATTTTCATGTTTTCCTCAAACT	1646
QY	98	-----	98

DB	1647	CATGAAGATGAATTAAGATTATTTTGCCTTCGTAATGCMAAACAGGTTTTTCATCTG	1706
QY	98	-----	98
DB	1707	TATTAATACATCATTCATTTTAAGAGACAAATATAACGGTCCAAGTGTGACTACT	1766
QY	98	-----	98
DB	1767	GCCCAAGATTTCAAGTGTCTATCAAGAAAGAAATTAACAATTACAGAGTGATGATG	1826
QY	99	-----ValasgluValArgArgLeuValTyr	107
DB	1827	GAAGTACTTGCGCGCATGATCTAATATCTGGGTGATGAAGCCAGAAAGCTGGGTAC	1886
QY	108	PheGluGluTyrThrAspSerProLeuGluHisIleuTyrValValSerTyrValAsn	127
DB	1887	TTTGAAGGACCAAGACCTCTCTTGGAGATCACCTGTAGTGACCAAGTATGCAAC	1946
QY	128	ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGln	147
DB	1947	CCTGAGAAAGTGTGAGGCTGACGACCGTGGCTACTACACTCCGCTGCTGACGCGG	2006
QY	148	HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu	167
DB	2007	CATTGTGACTCTTGTATAGTAAGTACAGCAACGAAAGAAATCCACATGTGTCTCC	2066
QY	168	TyrLysLeuSerSerProGluAsnAspProThrCysLysThrLysGluPheThrAlaThr	187
DB	2067	TACAACTCTCAAGTCTGAGGATGACCCAGTTCATAAACAAGAAATTTGGGCCACC	2126
QY	188	IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu	207
DB	2127	ATTTGGATTGAGAGGCTCCTCTCTGACTACACCCCTCGAATAATTTTCTTTGAA	2186
QY	208	SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuInProGly	227
DB	2187	AGTACTCTGATGATTTACACTGATGAAATGTTGTATTAAGCTTCACAGACCTCAAC	2246
QY	228	LysLysTyrProThrValLeuPheIleTyrGlyGlyProGluValGluLeuValAsnAsn	247
DB	2247	ANGAATACCCACTGTGTTTATTCATATATGAGTGTCCTCCAGGTGACCTGTGAACAT	2306
QY	248	ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal	267
DB	2307	CGGTTAAAGAGTCAAGATATTCGCGCTGAACACCTGCGCTCCCTGGTTATGCGTT	2366
QY	268	ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys	287
DB	2367	GTTGATGATGACAAACAGGGGATCCTGTGACCCAGGACCTTAAATTTGAAGCGCCTTAA	2426
QY	288	TyrLysMetGlyGluIleGluIleAspAspGluValGluGlyLeuGlnTyrLeuAlaSer	307
DB	2427	TATTAATGAGTCAATGAAATGATATGAGTGAAGGACTCCAGTACCTTAGCATCT	2486
QY	308	ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr	327
DB	2487	CAGTATGACTTCAATGACTGATGAGAGGACATCCACGCTGTGCTATGCTGCTAC	2546
QY	328	LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla	347
DB	2547	CTCTCCCTGATGCGATTAATGACAGGTCGAGTATCTCCGGGTGGCTATGCTGGGGCC	2606
QY	348	ProValThrLeuTyrIlePheTyrAspThrGlyTyrThr	360
DB	2607	CCAGTACCCCTGTGATCTTCTATGATACAGGATACAG	2645

RESULT 12
BC030688
LOCUS
DEFINITION
IMAGE:4822550, mRNA, complete cds.

BC030688 3130 bp mRNA linear PRI 30-MAY-2002
Homo sapiens, similar to dipeptidyl peptidase 8, clone MGC:26191

VERSION BC030688.1 GI:21265132
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 3130)
JOURNAL Strausberg, R.
Direct Submission
Submitted (24-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAC Plate: 33 Row: d Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 18450277.
FEATURES
source
Location/Qualifiers
1..3130
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26191 IMAGE:4822550"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
372..2867
/codon_start=1
/product="Similar to dipeptidyl peptidase 8"
/protein_id="AAH30688.1"
/db_xref="GI:21265133"
/translation="MAAMTEQLGVEIFPTADCEENISQDRPKLEPFYVERYSNQ
LKLILADTRKYHGMMAKAPHDPMFVKRNDPDPHSRIRIYILMSGENRENTLFSBI
PKTINRAAVLMLSMKPLDLFOATLDYGMTSREELIREKRIGTVGIASDYHQSG
TFLFOAGSGIYHVKDGGPOGFTOOPLPBNLVETSCHIRMDPLCPADPMIATISN
DIWISNIYREERRLTYVHNEILAMMEDASAGVATFVCOEEDRSGVWMCARATT
PSGKILIRIYRENDSEVRIHYTSPMLSTRASDRFRYFKGTANPKYTFKRSIMT
DAEGRIIDVIDKELIOPELTFEGVEITIAAGWTPBCKTAMSLILDRSRLQVLIS
PELFIPVEDVMERORLIESVPSVPLIYEBTDIWIHDIHFVFPQSHEEIEF
IPASECKTFRHLKYKITSILKESKYKSSGGLPAPSPKPIKEEIALITSGEWEVLGR
HGSNIQVDEVRRLVYFEGTKDSPLEHLLVYVYNPGBVRLDRGSHSCCISQD
FFISKSNQKPHCVSLYKLSPEDDPTCTKRFMTILISAGPLDPLPYRPEIFSES
TVGFTLYGMILYKHGDIORCKKPYVLPIGCGPOVLNNPFGVKVYKRLNTASLGV
VVVIDNRGSHRGKTEGAKYKVALIAGAPVTIMTFYDTGYTERKHGPDQNGY
LGSVVAQAEKFPSPBPNKLLHGLFDENVFATSTILSFLVAGKPYDILQIYQERH
SIRVPSGSEYELHLHYIQENLSRILAKVY"

BASE COUNT 955 a 626 c 717 g 832 t
ORIGIN

Alignment Scores:
Pred. No.: 4.55e-149 Length: 3130
Score: 1534.00 Matches: 309
Percent Similarity: 57.97% Conservative: 0
Best Local Similarity: 57.97% Mismatches: 0
Query Match: 79.36% Indels: 224
DB: 9 Gaps: 2

US-10-070-464-7 (1-360) x BC030688 (1-3130)

QY 1 GIUJLNASPALAArgSerAlaGlyValAlaThrPheValLeuGluGluGluPheAspArg 20
DB 1101 GAAGAGAGATCCAGATCAGCTGAGTGGTACCTTGTCTCCAGAAAGAAATTTGATPAG 1160
QY 21 TYSERGLYThrTyrTrpCysProLysAlaGluThrThrProSerGlyValLysIleLeu 40
DB 1161 TATTCGGCTATTTGGTGGTGTCCAAAGCTGAACACCTCCAGTGGTGTAAATTCCT 1220
QY 41 ArgIleLeuTyrGluGluLysAspGluSerGluValGluIleIleHisValThrSerPro 60
DB 1221 AGAATTCCTATGAGAAAGAAATGATGATCTGAGTGAATTAATTCATGTTACATCCCT 1280
QY 61 MetIleGluThrArgArgAlaAspSerPheAlaGlyTyrProLysThrGlyThrAlaAspPro 80
DB 1281 ATGTGGAAACAAAGAGGGGCAATTCCTGTTACCTPAAACAGGTACAGCAAAATCCCT 1340
QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle 98
DB 1341 AAGGTACCTTTTAAAGATGTCAGAAATTAATGATTTGATGCTGAAGAAAGATCATGATGTC 1400
QY 98 ----- 98
DB 1401 ATGATTAAGAACTAATTCACCTTTTGAATTCATTTGAAGAGTTGAATATATATGCGC 1460
QY 98 ----- 98
DB 1461 AGAGCTGATGATGCTCTGAGGAAATATGCTTGGTCCATCTCACTGATGCTGCCAG 1520
QY 98 ----- 98
DB 1521 ACTGCGCTACAGATAGTGTGATCTGACCTGATTAATTTATCCAGAGAAAGATGATGTT 1580
QY 98 ----- 98
DB 1581 ATGAAAGCAGAGACTCATATGATGATGCTGATTCGTGTGACGCCACTAATATATCTAT 1640
QY 98 ----- 98
DB 1641 GAAAGAAACAACAGACATCTGATTAATATCCATGACATCTTATGTTTCCCAAGT 1700
QY 98 ----- 98
DB 1701 CACGAGAGGAATTTGATTTATTTTGGCTGTAATGCAAAACAGGTTCCGTCATTTA 1760
QY 98 ----- 98
DB 1761 TACAAATTAATCATCTATTTTAAAGAAAGCAAAATATAAGCATCCAGTGGGCTGCT 1820
QY 98 ----- 98
DB 1829 GCTCCAAAGTATTTCAAGTGTCTCTATCCAAAGAGAGATAGCAATTACCACTGGTGAATGG 1880
QY 99 ----- ValAspGluValArgArgLeuValTyr 107
DB 1881 GAAGTTCCTGGCCGCGCATGATTAATATCAAGTGAATGATGAAGCAAGGCTGATAT 1940
QY 108 PheGluGlyThrLysAspSerProLeuGluHisHisIleuTyrValValSerTyrValAsn 127
DB 1941 TTGGAAGGACCAAAAGCTCCCTTTAGAGCAATCACCCTGACGATGACGTTCAGTAAT 2000
QY 128 ProGlyValValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
DB 2001 CCTGGAGAGGTGACAAAGCTGATGACGCTACGCTACATCTTGTCTCATAGTACAG 2060
QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGluLysAsnProHisCysValSerLeu 167
DB 2061 CACTGAGACTCTTTTATAGTATAGTATAGTAACAGAAAGATCACACTGCTGCTCCCT 2120
QY 168 TyrLysLeuSerSerProGlyLysAspProThrCysLysThrLysGluPheThrAlaThr 187
DB 2121 TACAGACTATCAAGTCTGAGATGACCACTTGCAGAAACAAAGAAATTTGGGCAACC 2180

QY 188 ILeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
| | | | |
Db 2181 ATTTGGATTACAGCAGGTCCTCTCCGACTATACCTCCAGAAATTTCTCTTTGAA 2240
| | | | |
QY 208 SerThrThyGlyPheThrLeuTyrGlyMetLeuTyrLeuPheHisAspLeuGlnProGly 227
| | | | |
Db 2241 AGTACTACTGAGATTTTCATTGATGGGATGCTCTTACAGGCTCATATCTACAGCCTTGA 2300
| | | | |
QY 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 247
| | | | |
Db 2301 AAGAAATATCCTACTGCTGCTGTTCAATATGGTGGTCCAGGTGAGTTGGGATTAAT 2360
| | | | |
QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
| | | | |
Db 2361 CGATTAAAGAGATCAAGATTTCCGCTGAATACCTTACCTCTAGCTTATGAGTT 2420
| | | | |
QY 268 ValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGlyValaPheLys 287
| | | | |
Db 2421 GTAGGATAGACACAGGGGATCTCTCCAGGGCTTAAATTTGAAAGCGCTTTTAA 2480
| | | | |
QY 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
| | | | |
Db 2481 TATTAATG----- 2489
| | | | |
QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheTyrGlyTyr 327
| | | | |
Db 2489 ----- 2489
| | | | |
QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
| | | | |
Db 2490 -----GTTGCTATTGCTGGGGCC 2507
| | | | |
QY 348 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
| | | | |
Db 2508 CCAGTCACCTGCTGATCTTCTATGATACAGATACAG 2546
| | | | |
RESULT 13
BD157001 2161 bp DNA linear PAT 17-JAN-2003
LOCUS BD157001
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD157001
VERSION BD157001.1 GI:27862759
KEYWORDS JP 2002191363-A/11844.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Ota,T., Isegai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 11844 09-JUL-2002;
HELIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/11844
PF 09-JUL-2002 JP 2000280990
PI TOSHIO OTA,TAKAO ISOgai,TETSUO NISHIKAWA,KOJI HAYASHI,KOJI
SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TETSUJI OTSUKI
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10, PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
FEATURES
source location/Qualifiers
FT CDS (16)..(1911).
1..2161
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 662 a 429 c 452 g 618 t
ORIGIN
Alignment Scores:
Pred. No.: 1,216-148 Length: 2161
Score: 1528.00 Matches: 308
Percent Similarity: 57.79% Conservative: 0
Best Local Similarity: 57.79% Mismatches: 1
Query Match: 79.05% Indels: 224
DB: 6 Gaps: 2
US-10-070-464-7 (1-360) x BD157001 (1-2161)
QY 1 GIUGAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluPheAspArg 20
| | | | |
Db 148 GAAAGAGATGCCAGATCAGCTGAGTCCGCTACTCTTTGTTCTCCAAAGAAATTTGATAGA 207
| | | | |
QY 21 TyrSerGlyTyrTyrTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleu 40
| | | | |
Db 208 TATTCGGCTATGGTGGTGGTCCAAAGCTGAAACAACTCCAGTGGGTAAATTTCTT 267
| | | | |
QY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
| | | | |
Db 268 AGAATTTCTATGAAAGAAATATGATGATCTGAGTGGAAATTTATGATTTACATCCCT 327
| | | | |
QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
| | | | |
Db 328 ATGTGGAACAAGAGGGGAGATCATTCCTGTTATCTTAAACAGGTACAGCAATCTCT 387
| | | | |
QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle 98
| | | | |
Db 388 AAGTCACCTTTTAAAGATGTCAGAAATATGATGATGTCGAAAGAAAGATCATAGATGTC 447
| | | | |
QY 98 ----- 98
| | | | |
Db 448 ATAGATTAAGAACTAATTCACCTTTTGAGATTCTATTGAAAGAGTTGAATATATTGCC 507
| | | | |
QY 98 ----- 98
| | | | |
Db 508 AGAGCTGATGAGACTCTGAGGAAATATATGCTTGTCATCTACTAGTCCTCCAG 567
| | | | |
QY 98 ----- 98
| | | | |
Db 568 ACTGCGCTACAGATAGTGTGATCTCACCTGATTTATTTATCCAGTAGAGATGATGTT 627
| | | | |
QY 98 ----- 98
| | | | |
Db 628 ATGGAAGCAGAGACTCATTTAGTCAGTGCCTGATTTCTGTGACGCCACTAATTATCTAT 687
| | | | |
QY 98 ----- 98
| | | | |
Db 688 GAAAGAAACAAGACATCTGATTAATATCATGACATCTTCATGTTTTTCCCAAGT 747
| | | | |
QY 98 ----- 98
| | | | |
Db 748 CAGGAAGAAATGAGTTATTTTGGCTCTGAAAGCAAGATTTCCGTCATTTA 807
| | | | |
QY 98 ----- 98
| | | | |
Db 808 TACAAATTTACATCTATTTTTAAAGAAACAAATATTAACGATCCAGTGGTGGCTCCCT 867
| | | | |
QY 98 ----- 98
| | | | |
Db 868 GCTCCAGTATTTCAAGTGTCTATCAAGAGAGATAGCAATTAACAGTGGTGAATGG 927
| | | | |
QY 99 -----ValAspGluValArgArgLeuValTyr 107
| | | | |
Db 928 GAAGTCTTGGCCGCGCATGATCTAATATCCAAAGTTGATGAGATGAGAGGCTGATATAT 987
| | | | |
QY 108 PheGluGlyThrLysAspSerProLeuGluHisIleLysTyrValValSerTyrValaAsn 127
| | | | |
Db 988 TTTGAAGCACCAAGATCTCCCTTTAAGCATCCCTGTACTGATGCTTACGTTACGTAAT 1047
| | | | |

QY 128 ProGlyValThrArgLeuThrAspArgGlyTyrSerHisSerCysSerGln 147
 Db 1048 CCTGAGAGGTACCAAGGCTGACGACCTGACCTCACTTCTGCGATCGCTAG 1107
 QY 148 HisCysAspPhePheIleSerIleTyrSerAangInLysAspProHisCysValSerLeu 167
 Db 1108 CACTGGACTTTTATTAAGTAAAGTAACTAACAGAAATCCACAGCTGTGTCCCTT 1167
 QY 168 TyrIleLeuSerSerProGlnAspAspProThrCysIleThrIleGluPheTyrAlaThr 187
 Db 1168 TACAGACTATCAAGCTCCTGAAGATGACCCACCTGCAAAACAAAGAAATTTGGGACACC 1227
 QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGlnIlePheSerPheGlu 207
 Db 1228 ATTTGGATTCAAGAGGCTCTCTCTGACATACCTCCCAAGAAATTTCTCTTTGAA 1287
 QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
 Db 1288 AGTACTACTGAGTTTACATTGATGAGGATGCTCTCAAGCCTCATGATCTACAGCCTGGA 1347
 QY 228 LysIleTyrProThrValIleuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
 Db 1348 AAGAAATATCCATCACTGCTGCTGATATATGATGCTGCTCAGCTGCAATGATATAT 1407
 QY 248 ArgPheLysGlyValIleTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 Db 1408 CGGTTTAAAGAGCTCAAGATATTCGCTTGAAATACCTTACCTCTCTAGATTATGTGTT 1467
 QY 268 ValValIleLeuAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
 Db 1468 GATAGATATAGACAAACAGGGGATCTGTCACAGGGCTTAAATTTGAAGGCCCTTTTAA 1527
 QY 288 TyrIleMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
 Db 1528 TATAAAATG----- 1536
 QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 327
 Db 1536 ----- 1536
 QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
 Db 1537 -----GTTGCTATTGCTGCGGCC 1554
 QY 348 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
 Db 1555 CCAAGTACTCTGAGATCTTATGATACAGGATACAGC 1593
 RESULT 14
 AK027826 2161 bp mRNA linear PRI 01-AUG-2002
 LOCUS Homo sapiens cDNA FLJ14920 fis clone PLACE1007416, weakly similar
 DEFINITION to D18PPTIDYL PERIPIIDSE IV (EC 3.4.14.5).
 ACCESSION AK027826
 VERSION AK027826.1 GI:14042789
 KEYWORDS oligo cloning; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1
 AUTHORS Iisogai,T., Ota,T., Hayaashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
 Wagaitsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
 Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
 Ishii,S., Kawai,Y., Salto,K., Yamamoto,U., Wakamatsu,A.,
 Nakamura,Y., Nagahashi,K., Masuo,Y., Nimomiyu,K. and Iwayanagi,T.
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2161)
 AUTHORS Iisogai,T. and Otsuki,T.
 TITLE Direct Submission

JOURNAL

Submitted (10-MAY-2001) Takao Iisogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 E-mail:genomics@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology; cDNA library construction,
 5'- & 3'-end one pass sequencing and clone selection; Helix
 Research Institute (supported by Japan Key Technology Center etc.)
 and Department of Virology, Institute of Medical Science,
 University of Tokyo.

FEATURES

source

1..2161
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="PLACE1007416"
 /tissue_type="Placenta"
 /clone_id="PLACE1"
 /note="cloning vector: pME18FLJ3"
 16..1914
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB55395.1"
 /db_xref="GI:14042790"

CDS

translation="MDPKLCPADPDWIAFIHSNDIWSINIVTRERRRLTYVNEILAM
 EEDARSAGVATPVQEDRRYSQYWCWKAKATTPEGKILILYEENDESVEIITHY
 SPMLSTRADSFRIPIKGTANPKYTFKSEIMIDBSKILIDVIDELIOPETILEGV
 EYIARAGMTPEGKYAWSILLDRSQRLQVILISPELFIPEDDWNERQLIESVDSV
 PLPIIYETETIDWINIHDIFHFQSHSEEEEFIFASECKTGRILYVITSLKESKY
 KRSGGLPAPSDPFKCFIEKRIATSGSEVEYRGNSNIOVEERLIVYETGKDSFLE
 HILVVSYVNGEYTRLLDRSYSHSCCISQCHDPFISIKYSNOKNHCYSLVLSQSPD
 DPCKTERFMAITLIDASRLPDYTPPEIFRSESTGPTLYKMLYRPHLOPKKPTV
 LFIYGGPQVLVNNKPFQVKYFRMLTSLQVILVVVIVNNRSCHEGLKFEAGFKYKM
 AIAQAPVLMIFDYGTERTYMGHDQEQSYLQSVVMQAEKFPSENRLLIAGFL
 DENHFATSTILSFLVYAGKPYDLQIYPORRHSIRVPSGEHYELHLHYQENLGS
 RIAALKVY"

BASE COUNT 662 a 429 c 452 g 618 t
 ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2161
 Score: 1528.00 Matches: 308
 Percent Similarity: 57.79% Mismatches: 1
 Best Local Similarity: 57.79% Indels: 224
 Query Match: 79.05% Gaps: 2
 DB:

US-10-070-464-7 (1-360) x AK027826 (1-2161)

QY 1 GIUGIUAAPPAIARYSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
 Db 148 GAAGAGAGATGCGAGATGAGTCCCTACCTTGTCTCCCAAGAAATTTATATGAA 207
 QY 21 TyrSerGlyTyrTyrTyrCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
 Db 208 TATTGTGCTATTGTGTGTGTGCCAAACCTGAACAACTCCACAGTGGTGAATTTCTT 267
 QY 41 ArgIleLeuTyrGluGlnAsnAspGluSerGluValGluIleHisValThrSerPro 60
 Db 268 AGAATTTATATGAGAAAGAAAGATGAATCTGAGGGAGAAATTTATCATGATACATCCCT 327
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
 Db 328 ATGTTGAGAAACAGAGAGGCGATTCATTCCTTATCTTAACACAGTACAGCAATCTCT 387
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
 Db 388 AAGATCACTTTTAAGATGTCAGAAATATGATGATGCTGGAAGAGATCATGATGTC 447
 QY 98 ----- 98
 Db 448 ATGATTAAGAGACTTAATTCACCTTTTGAATGATCTATTGTAAGAGAGTTGAATATATTGCC 507


```

QY 98 ----- 98
Db 508 AGAGCTGGATGACTCCTGAGGAAAAATATGCTTGCCATCCTACTAGATCGCTCCAG 567
QY 98 ----- 98
Db 568 ACTCGCTACAGATAGTGTGATCTCAGCTGAATTAATTTATCCAGTAGAAGATGATGTT 627
QY 98 ----- 98
Db 628 ATGAAAAGCAGAGACTCATTGATGCTGATTCGTGTGACGCCACTAATATATCTAT 687
QY 98 ----- 98
Db 688 GAAAGAAACAAGACATCTGGATTAATATCCATGACATCTTTCATGTTTTCCCAAAGT 747
QY 98 ----- 98
Db 748 CACGAAGAGAAATGAGTTATTTTTCCTCTGAATGCAAAAAGGTTCCGTCATTTA 807
QY 98 ----- 98
Db 808 TACAAATTTACATCTATTTTAAAGAAAGCAATATTAACGATCCAGTGGGCTGCCT 867
QY 98 ----- 98
Db 868 GCTCCAGTGATTTCAAGTGTCTTATCAAAGAGATGCAATTAACAGTGTGAAATG 927
QY 99 ----- ValaspGluValArgArgLeuValTyr 107
Db 928 GAAGTCTTGCGCGGATGATGCTAATATTCAGATGTTGATTAATCAGAGGCTGGTATAT 987
QY 108 PheGluGlyThrTyrAspSerProLeuGluHisLeuTyrValValSerTyrValAsn 127
Db 988 TTTGAAAGGACCAAAAGCTCCCTTAAGAGCATCACCTGTAAGTACAGTACGTAAT 1047
QY 128 ProGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysHisSerArg 147
Db 1048 CCTGGAGAGGTGACAGGCTGACTGACCGTACCTACTCAATCTTGCTGCACACAGCTCAG 1107
QY 148 HisCysAspPhePheIleSerTyrSerAsnGluLeuAsnProHisCysValSerLeu 167
Db 1108 CACTGTGACTTCTTTTAAGTATAGTATACCAAGATCAACACTGTGTCTCTT 1167
QY 168 TyrTyrLeuSerSerProGluAspAspProThrCysLeuThrTyrGluPheTyrPalat 187
Db 1168 TACAACTATCAAGTCTGAGATGACCAAACTTGCAAAAAGAAATTTTGCGCACCC 1227
QY 188 IleLeuAspSerIleGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db 1228 ATTTTGATTTCAAGAGGTCCTCTCTGACTATATCTCCAGAAAATTTTCTTTTGA 1287
QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrTyrProHisAspLeuGluProGly 227
Db 1288 AGTACTACTGATTTACATTGATGATGGAGATGCTCTACAGCTCATATCTACAGCTCGA 1347
QY 228 TyrTyrTyrProThrTyrIleuPheIleTyrGlyGlyProGluValGluLeuValAsn 247
Db 1348 AAGAAATATCCACTGTGTCTGTCATATATGTTGGTCCCTCAGGTGACAGTGTGATAT 1407
QY 248 ArgPheLeuGlyValIleTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValIle 267
Db 1408 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACGCTCTTACGTTATGCGT 1467
QY 268 ValIleValIleAspAsnArgIleSerCysHisSerGlyLeuLeuPheGluGlyAlaPheTyr 287
Db 1468 GTAAGCATAGACAACAGGGATCTCTGTCCAGGAGGCTTAATTTGAAGCGGCTTTAA 1527
QY 288 TyrTyrMetGlyGluIleGluIleAspArgIleValGluGlyLeuGlnTyrLeuAlaSer 307
Db 1528 TATTAATATG----- 1536
QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyTyr 327

```

```

Db 1536 ----- 1536
QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleGlyAla 347
Db 1537 ----- GTTCTATTGCTGGGCCC 1554
QY 348 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
Db 1585 CCAGTCACTCTGTGATCTTCTATGATATACGATATACG 1593

RESULT 15
AX338497 2510 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 16 from Patent WO0183775.
DEFINITION AX338497
ACCESSION AX338497
VERSION 1
KEYWORDS AX338497.1 GI:18128919
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Deleane, A.M., Lal, P., Hafalia, A., Patterson, C., Walla, N.K.,
Keamey, L., Tribouley, C.M., Khan, F.A., Yao, M.G., Baughn, M.R.,
Azimzai, Y., Elliott, V.S., Nguyen, D.B., Gandhi, A.R., Yang, J.,
Hernandez, R., Policky, J.L., Lu, D.A., Reddy, R. and Tang, Y.T.
Proteases
Patent: WO 0183775-A 16 08-NOV-2001;
Incyte Genomics, Inc. (US)
LOCATION/Qualifiers
SOURCE
1. 2510
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 376067CB1"

BASE COUNT 777 a 494 c 527 g 712 t
ORIGIN

Alignment Scores:
Pred. No.: 6,12e-148 Length: 2510
Score: 1522.00 Matches: 309
Percent Similarity: 58.08 Conservative: 0
Best Local Similarity: 58.08 Mismatches: 223
Query Match: 78.74% Indels: 2
DB: 6 Gaps: 2

US-10-070-464-7 (1,360) x AX338497 (1-2510)
QY 1 GluIleAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGluIleAspArg 20
Db 609 GAAAGAGATGCCAGATCACTGAGTGGCTACCTTTGTTCTCCAAAGAAATTTGATAGA 668
QY 21 TyrSerGlyTyrTyrProCysProTyrAlaGluThrThrProSerGlyGlyTyrIleu 40
Db 669 TATTCGGCTATTTGGTGGTCCAAAGCTGAAACAACTCCAGTGTGTAAATTTCTT 728
QY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
Db 729 AGAATCTATAGAGAAATGATGATCTGAGTGGAAATTTATCATGTTACATCTCT 788
QY 61 MetLeuGluThrArg-ArgAlaAspSerPheArgTyrProTyrThrGlyThrAlaAsn 80
Db 789 ATGTGGAACAAGGAGGAGATTCATTTCCGTTATCTTAAACAGTACAGCAATCC 848
QY 80 GlyValThrPheTyrMetSerGluIleMetIleAspAlaGluGlyArgIle----- 98
Db 849 TAAAGTCACTTTTAAAGATGTCAGAAATATATGATGATGTCGAAGAGATTCATATAT 908
QY 98 ----- 98
Db 909 CATAGATAAGAACTAATTCAACCTTTTGAGATTCTATTGTAAGAGATTGAATATATTC 968

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:14:56 ; Search time 264.512 Seconds

(without alignments)
3673.932 Million cell updates/sec

Title: US-10-070-464-7

Perfect score: 1933

Sequence: 1 EEDARAGAVATFVLOEPRD.....RNAAGAPVTIMIFDTGTT 360

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODER=frame+.p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10070464/runat.15102003.113552.24799/bpp.query.fasta_1.2652
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=mg -MINMATCH=0.1 -LOOPEXT=0
-LIST=45 -DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10070464@cgn2.1.0@runat.15102003.113552.24799 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: N Geneseq 19Jun03: *
2: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1980.DAT: *
3: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1981.DAT: *
4: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1982.DAT: *
5: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1983.DAT: *
6: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1984.DAT: *
7: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1985.DAT: *
8: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1986.DAT: *
9: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1987.DAT: *
10: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1988.DAT: *
11: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1989.DAT: *
12: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1990.DAT: *
13: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1991.DAT: *
14: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1992.DAT: *
15: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1993.DAT: *
16: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1994.DAT: *
17: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1995.DAT: *
18: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1996.DAT: *
19: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1997.DAT: *
20: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1998.DAT: *
21: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA1999.DAT: *
22: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA2000.DAT: *
23: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA2001A.DAT: *
24: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA2001B.DAT: *
25: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA2002.DAT: *
26: /SIDSI/gcgdata/geneseg/genesegn-emb1/NA2003.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1933	100.0	1083	22	Nucleotide sequenc
2	1880	97.3	4309	22	CDNA encoding huma
3	1836.5	95.0	2643	24	Coding sequence of
4	1836.5	95.0	2649	25	CDNA encoding huma
5	1836.5	95.0	2671	24	CDNA encoding huma
6	1836.5	95.0	3106	24	CDNA encoding huma
7	1836.5	95.0	3120	22	Human protease PR
8	1836.5	95.0	3120	24	Nucleotide sequenc
9	1836.5	95.0	3143	24	Human dipeptidyl p
10	1816.5	94.0	4829	24	CDNA encoding 2195
11	1528	79.0	2151	22	CDNA encoding huma
12	1522	78.7	2510	22	Human CDNA sequenc
13	1522	78.7	2510	24	Human CDNA sequenc
14	1521.5	78.7	4676	24	CDNA encoding huma
15	1429	73.9	1821	24	CDNA encoding huma
16	1254.5	64.9	2668	24	Dipeptidyl peptida
17	1254.5	64.9	2668	24	Novel human coding
18	1229	63.6	4523	24	Novel human coding
19	1220.5	63.1	2495	24	CDNA encoding huma
20	1220.5	63.1	2617	24	Human dipeptidyl p
21	1220.5	63.1	3024	24	CDNA encoding huma
22	1220.5	63.1	3716	24	Human dipeptidyl p
23	1220.5	63.1	4219	24	Human PMW encodin
24	1220.5	63.1	4302	24	CDNA encoding huma
25	1204.5	62.3	2751	24	CDNA encoding huma
26	1204.5	62.3	3287	24	Murine dipeptidyl
27	1201	62.1	4180	24	Alternative versio
28	1201	62.1	4263	24	CDNA encoding huma
29	1183.5	61.2	3262	22	CDNA encoding huma
30	1169.5	60.5	2801	22	Human polynucleoti
31	1169.5	60.5	2982	22	Human polynucleoti
32	1132.5	58.6	4076	24	Human polynucleoti
33	1132.5	58.6	4159	24	CDNA encoding huma
34	1113	57.6	4037	24	CDNA encoding huma
35	1113	57.6	4120	24	CDNA encoding huma
36	1086.5	56.2	1197	22	CDNA encoding huma
37	984	50.9	2952	24	Nucleotide sequenc
38	984	50.9	3047	24	DNA encoding huma
39	944	48.8	2027	21	Human ORFX OR2692
40	860.5	44.5	1669	22	Nucleotide sequenc
41	722	37.4	3713	23	Drosophila melanog
42	722	37.4	3783	23	Drosophila melanog
43	654.5	33.9	2461	21	Human ORFX ORP1390
44	558	28.9	6225	23	Drosophila melanog
45	558	28.9	6228	23	Drosophila melanog

ALIGNMENTS

RESULT 1
ID AAC85697 standard; cDNA; 1083 BP.
AAC85697;
29-JUN-2001 (first entry)
Nucleotide sequence encoding human DPP8 244Glu-341Ile+515Val-776Thr.
Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhea;
growth hormone deficiency; glucose level; mucosal regeneration;
non-insulin dependent diabetes mellitus; glucose intolerance;
immunosuppression; ss.
Homo sapiens.
OS

XX Key Location/Qualifiers
 FH CDS 2..1081
 FT /tag= a
 FT /product= "Human DPP8 244Glu-341Ile+51SVal-776Thr"
 FT /note= "No stop codon given"
 XX MO200119866-A1.
 XX
 XX 22-MAR-2001.
 XX
 XX 11-SEP-2000; 2000MO-AU01085.
 XX
 XX 10-SEP-1999; 99AU-0002762.
 XX 18-FEB-2000; 2000AU-0005709.
 XX
 XX (UNSY) UNIV SYDNEY.
 XX
 XX Abbott CA, Gorell MD;
 XX
 XX MPI; 2001-281520/29.
 XX P-PSDB; AAB47150.
 XX
 XX New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
 PT substrates, identifying inhibitors of DPP8 catalytic activity which
 PT have therapeutic uses, and for detecting activated T cells -
 XX
 XX Claim 21; Page 75-76; 78pp; English.
 XX
 CC The sequences given in AAC85695-97 encode fragments of human
 CC dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for
 CC H-Ala-Pro-pNA, H-Gly-Pro-pNA and H-Arg-Pro-pNA. Therefore, it is a
 CC prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable
 CC of hydrolysing the peptide bond C-terminal to Pro in each of these
 CC compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for
 CC cleaving a substrate, and for detecting an activated T cell which
 CC involves measuring the level of DPP8 gene expression in a T cell. The
 CC level of DPP8 expression is detected by detecting the amount of DPP8
 CC RNA in the cell. It is also useful for identifying a molecule capable
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules
 CC identified as inhibiting DPP8 catalytic activity may be useful for
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
 CC in non-insulin dependent diabetes mellitus and other disorders
 CC involving glucose intolerance, enhancing mucosal regeneration and
 CC as immunosuppressants.
 XX
 XX Sequence 1083 BP; 312 A; 213 C; 247 G; 311 T; 0 other;
 XX
 XX
 XX Alignment Scores:
 XX Pred. No.: 1,46e-216 Length: 1083
 XX Score: 1933.00 Matches: 360
 XX Percent Similarity: 100.00% Conservative: 0
 XX Best Local Similarity: 100.00% Mismatches: 0
 XX Query Match: 100.00% Indels: 0
 XX DB: 22 Gaps: 0
 XX
 XX US-10-070-464-7 (1-360) x AAC85697 (1-1083)
 XX
 QY 1 GIUGLUSPAAARLARGSERALAGLYVALAIAATHRPHEVALLEUNGINGUINPHEAPARG 20
 DB 2 GAAGAAGATGCCAGATCAGCTGAGTGGCTACCTTGCTCCMAAAGAAATTTGATAGA 61
 QY 21 TYRSEGLTYRTTPTTPCSPPOLYSALAGLUTHTHPROSEGLVGLVLSILEU 40
 DB 62 TATTCGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCAGTGGTAAATTTCTT 121
 QY 41 ARGILEUITYGILUGLUSNAPSGIUSERGIVAGLULLEHISVALTHRSERPRO 60
 DB 122 AGAATTCATATGAAATAATGATGATCGAGTGGAAATTTATCATGTTACATCCCT 181
 QY 61 METLEUGLUTHRARGAAGLAASERPEARGTYPRULYSTRHRGLYTHRAAENPRO 80
 DB 182 AGTTTGAAACAGAGGCGCAGATTCATTCGTTATTCCTPAAACAGGTACAGCAAAATCCT 241

QY 81 LYSVALTHRPHELYSMESERGIULEMETILEASPLAIGLGLYVARGILEILEVALSP 100
 DB 248 AAAGTACTTTTAAAGTGTCCAGAAAATAAGATTGAGCTGAAAGAAAGATCATAGTTGAT 301
 QY 101 GIUVALARGARLEUVALTYRPHGUGLYTHRLYSASPSERPROLEUGNHSISLEU 120
 DB 302 GAAGTCAGAAAGCTGGTATATTTTGAAGGACCAACMAAGATCCCTTTAAGCATCACTG 361
 QY 121 TYRVALVALSERTYRVALANSPROGLYUVALTHRARGLEUTHRASPARGLTYRSE 140
 DB 362 TAGGTAGTCAATTACCTTAATCCTGAGAGGTACAAAGCTGACTGACCGTGGCTACTCA 421
 QY 141 HISSECYCYEALLSEERGLNHSICYSAPPHEPHEILSESLTYRSEARNGNLVS 160
 DB 422 CATTCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 481
 QY 161 ASPPROHISCYSEVALSERLEUTYRYSLEUSERSESPROGLUSAPSPROTHRYSLYS 180
 DB 482 AATCCACACTGTGTGCTCCCTTTACAAAGCATATCAAGTCCGMAAGATACCACTTCGMA 541
 QY 181 THRYSGLUPHETIPALATHRIILEUASPSERALAGLYPROLEUPROASPTYRTHPR 200
 DB 542 ACAAGGAATTTTGGGCCACCAATTTGATGATGATGATGATGATGATGATGATGATGAT 601
 QY 201 PROGLIULPHESESPHEGLUSERTHRPHGLYPHEPHELEUTHRYGLYMETLEUTHRY 220
 DB 602 CCAGAAATTTTCTTTTAAAGTACTGATGATGATGATGATGATGATGATGATGATGATGAT 661
 QY 221 PROHISASPLEUGLNPROGLYLSLYSTYRPROTHRVALLEUPHEILETYRGLYGLYPR 240
 DB 662 CCTCAGATCATACAGCTGGAAGAAATTCCTACTGTCTGCTTCAATATGATGATGATGAT 721
 QY 241 GINVALGILEUVALSNAPSGIUSERGIVAGLYVALYSYRPHARGLEUANTHRLU 260
 DB 722 CAGGTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 781
 QY 261 ALASERLEUGLYTYRVALVALVALVALLIENASPARGLYSERCYSHASARGLYLEU 280
 DB 782 GGCTCTCTAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 841
 QY 288 LYSPLHGLUGLYALAPHELYSTYRYSMEGLYGLNILEGLIILEASPSAPGINVALGL 300
 DB 842 AATTTGAAAGCGCCCTTAAATATTAAGGCTCAATGAAATTAACATCACTGATCGGAA 961
 QY 301 GILEUGINTYRLEUALASERARGTYRASPHEILEASPLEUASPARVALGIYILEHS 320
 DB 902 GCACTCCATATCTAGCTTCTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 961
 QY 321 GLYTRPSERTYRGLYGLYTYRLEUSERLEUMELALILEUMEGLINRGSEARPIEPHE 340
 DB 962 GGCTGGTCTTATGAGAGATPACCTTCCCTGATGATGATGATGATGATGATGATGATGAT 1021
 QY 341 ARGVALAILEALAGLYALAPROVALTHRLEUTRPIEPHELYRASPTRGLYTYRTH 360
 DB 1022 AGGGTTGCTATTGCTGGGCCCAAGTCACTGCTGATCTTCTATATACAGATATCACG 1081
 XX
 XX RESULT 2
 XX ABR83328
 XX ID ABR83328 standard; cDNA; 4309 BP.
 XX
 XX ABR83328;
 XX
 XX 12-AUG-2002 (first entry)
 XX
 XX
 XX cDNA encoding human DPP-1 splice variant #4.
 XX
 XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
 XX DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 XX diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 XX heart failure; hypertension; urinary retention; osteoporosis; cancer;
 XX ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 XX dyskinesia; reproductive disorder; inflammatory disorder;

KM metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
XX 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
XX
DR P-PSDB; ABG61597.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX
XX
PS Disclosure; Page 67-68; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyslexias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPPR proteins.
XX
SQ Sequence 4309 BP; 1304 A; 790 C; 907 G; 1308 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 1.95e-209 Length: 4309
Score: 1880.00 Matches: 358
Percent Similarity: 98.90% Conservative: 0
Best Local Similarity: 98.90% Mismatches: 2
Query Match: 97.26% Indels: 3
DB: 24 Gaps: 0
US-10-070-464-7 (1-360) x ABK83328 (1-4309)
QY 1 GluGluASPAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
Db 943 GAAGAGATGCCAGATCGAGTGCCTGCTTGTCTTCCAGAGAAATTATATAGA 1002
QY 21 TyrSerGlyTyrTyrPyrCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
Db 1003 TATCTGCTATGTGCTGCTGCCAAAGCTGMAAACCTCCAGGCGTGAATAATCTT 1062
QY 41 ArgIleLeuTyrGlnGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
Db 1063 AGAATCTTATATGAAGAAATGATGATCTGAGGGGAAATATATCAATGATCAATCCCT 1122
QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
Db 1123 ATGTTGAAACAAGAGGCGAGATTCATTCGTTATCTTAAACAGATACGCAATCTT 1182
QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAsp 100
Db 1183 AAAGTCACTTTTAAGATGTCAGAAATATGATGATGCTGAAGAGAAATCCAA-GTTGAT 1241

QY 101 GluValAlaGArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisIleLeu 120
Db 1242 GAAGTCAGAAAGCTGTATATTTTGAAGGCACCAAGATCCCTTTAGAGATCACTCG 1301
QY 121 TyrValIleSerTyrValAsnProGluValThrArgLeuThrAspArgLysTyrSer 140
Db 1302 TACGTAGTCAGTTAGCTAAATCTCTGAGAGGTGACAGGCTGACGAGCTGACTCA 1361
QY 141 HisSerCysSileSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 160
Db 1382 CATTCTTGTCATCAGTCAGACCTGACTCTTTATATAGATAGATATGTAACCAAG 1421
QY 161 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 180
Db 1422 AATCCACACTGTGTCTCCCTTACAAAGTATCAAGTCTCGAAGATGACCACTTGCAA 1481
QY 181 ThrLysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 200
Db 1482 ACAAAAGAAATTTGGCCACCATTTTGGATTTAGAGAGGTCTCTTCTGACTATACTCT 1541
QY 201 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 220
Db 1542 CCAGAAATTTCTCTTTGAAAGTACTAGATTTACATTTATGGGATGCTCTACAG 1601
QY 221 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyLys-P 240
Db 1602 CCTCATGATCTACACCTGGAAGAAATATCTTACTGCTGCTTATATATATGATGCTCTC 1661
QY 240 roGlnValGlnLeuValAsnAsnArgPheLysGlyValIleTyrPheArgLeuAsnThrL 260
Db 1662 CTCAGGTGCACTTGTTGATTAATCGTTTAAAGAGTCAAGATATTCCTGTAATACCC 1721
QY 260 euAlaSerLeuGlyTyrValValValIleAspAsnArgLysSerCysHisArgGlyL 280
Db 1722 TAGCTCTCTAGTTATGATGTTGATGATAGATAGACACAGGGAGATCCTGTACCGAGGC 1781
QY 280 euLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValG 300
Db 1782 TTAAATTTGAAGGCCCTTTAAATTAATAAAGGGTCAATATGAATTCAGATCAGTGG 1841
QY 300 LysIleLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValIleHis 320
Db 1842 AAGGATCTCAATATATATCTTACTCTCTCATATGATTTATGACTTATGATGCTGGGGATCC 1901
QY 320 LysGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIleP 340
Db 1902 ACGGCTGCTCTATGAGAGATACCTCTCCCTGATGCGATTTATGACAGGTCAGATATCT 1961
QY 340 heArgValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrT 360
Db 1962 TCAGGGTTGCTATGCTGCGGCCCAATCCTGCTGTGATCTTCTATGATACAGGATACA 2021
QY 360 hr 360
Db 2022 CG 2023
RESULT 3
AAH99935
ID AAH99935 standard; cDNA; 2643 BP.
XX
AC AAH99935;
XX
DT 12-APR-2002 (first entry)
XX
DE Coding sequence of 21953 human prollyl oligopeptidase.
KW 21953 prollyl oligopeptidase; antibody; proline; endopeptidase;
KW cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
KW diabetes mellitus; arthritic; multiple sclerosis; asthma;
KW Grave's disease; neuronal disorder; demyelinating disease; ss.
XX

OS Homo sapiens.
XX
XX WO200179473-A2.
XX
XX 25-OCT-2001.
XX
XX 11-APR-2001, 2001WO-US40483.
XX
XX 18-APR-2000, 2000US-197508P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Meyers RA, Williamson M;
XX
XX WPI, 2002-034353/04.
XX
XX P-PSDB; AAG78415.
XX
XX New polypeptides 21953, member of human prollyl oligopeptidase family,
PT useful as diagnostic targets and therapeutic agents for controlling
PT cancer, lymphoma and leukemia -
XX
XX
XX Claim 2; Page: 121pp; English.
XX
XX This invention relates to an isolated 21953 human prollyl
CC oligopeptidase, which is cytosolic, antidiabetic, antiarthritic,
CC neuroprotective, antithyroid, dermatological, antipsoriatic,
CC antiaesthetic, ophthalmological, antiinflammatory, nootropic,
CC antiparkinsonian, anticonvulsant, gynecological, vasotropic,
CC antitubercular, cardiatic, antihypertensive, anorectic and
CC metabolic in its action. Uses include gene therapy, expression or
CC activity of 21953 protein modulator, it is useful for identifying a
CC compound which binds to it and can be used in preventing, treating
CC or detecting a cellular proliferative or differentiative disorder.
CC The 21953 molecules can act as novel diagnostic targets and therapeutic
CC agents for controlling disorders associated with the aberrant activity
CC or degradation of peptide hormones e.g., disorders associated with cell
CC differentiation and proliferation such as cancer, immune function,
CC reproductive, neurological and cardiovascular function. The 21953
CC molecules are thus useful for treating and preventing cellular
CC proliferative and differentiative disorders, haematopoietic neoplastic
CC disorders, immune disorders such as autoimmune diseases, diabetes
CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC neuronal disorders, demyelinating diseases, vascular disorders and
CC metabolism or pain disorders. This sequence represents the cDNA
CC encoding sequence of 21953 human prollyl oligopeptidase. This
CC sequence represents the coding sequence of 21953, being the sequence
CC in between the start and the stop codon of the sequence represented in
CC AAH99934.
XX
XX Sequence 2643 BP; 800 A; 514 C; 585 G; 744 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 1,19e-204 Length: 2643
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 24 Gaps: 1
US-10-070-464-7 (1-360) x AAH99935 (1-2643)
QY 1 GIUGIUSPALAARXSERIAGIYVALAATHRPHEVALLEUGINGIUNIPHEASPAR 20
DB 727 GAAGAAGATGCCAGATCAAGTGGAGTCTTGTCTCCAGAAAGAAATTTGATAGA 786
QY 21 TYRSEGIYTYTITPTCYSPROLYSALAGIUTNRTHPRSEGIYGLIYLSILEU 40
DB 787 TATTTGGCTATTGGTGGTGTCCAAAGCTGAACCACTCCAGTGGTGTAAATTTCTT 846
QY 41 ARGILEUITYRGIGIUNASASPGLUSERGIUVALGILIEIISHISVALTHRSEPR 60
DB 847 AGAATTCATATGAGAAAGAAATGATGATGTGAGGTGAAATTAATTCATTCATCCCT 906

QY 61 MELLEUGIUTHRAXARGHIAASPSERPHENGYTYRPROLYETHRGILYHRAIAENPRO 80
DB 907 ATGTGGAAACAAGAGGGGACATTCCTCCGTATCTTAACAAGGTACAGAAATTCCT 966
QY 81 LYSVALTHRPHELYMESERGIULEMETILEASPAIAGIUGIYARGILEILE----- 98
DB 967 AAGTCACCTTTTAAGTATGCAAAATAATGATTGATGCTGAAGAGAGATCATAGATGC 1026
QY 98 ----- 98
DB 1027 ATAGATAAGACATAATTCACCTTTGAGATTCTATTGGAAGAGTGAATATATTGCC 1086
QY 98 ----- 98
DB 1087 AAGCTGATGAGATCTCGAGGAAAATATGCTTGTCATCTACTAGATGCTCCAG 1146
QY 98 ----- 98
DB 1147 ACTGCGCTGCAGATAGTGTGATTCACCTGAATTTATTATCCAGTAGAAGATGATT 1206
QY 98 ----- 98
DB 1207 ATGAAAGGACAGACTCATGTGACGTGCTGATTCTGTGACGCCACTAATATCTAT 1266
QY 98 ----- 98
DB 1267 GAAGAAACACAGACATCTGATTAATATCATACATCTTTCATGTTTCCCAAGT 1326
QY 98 ----- 98
DB 1327 CACGAAGAGAAATGAGTTATTTTGGCTGTGAATGCAAAACAGTTCCGTCTTAA 1386
QY 98 ----- 98
DB 1387 TACAAAATTACATCTATTATTAAGAAAGCAATATTAACGATCCAGTGTGCTGCT 1446
QY 98 ----- 98
DB 1447 GCTCCAAAGATTCAAGTCTCTTCAAAAGAGATAGCAATTAACAGTGTGAATGG 1506
QY 99 -----ValASPGLUVALARGRGLUVALTYR 107
DB 1507 GAAGTCTCTGGCGGCGATGATCTAATATCCAAAGTTGATGAAGTGAAGGCGGTATAT 1566
QY 108 PHEGIUGIYTHLYASPSERPROLEUGIWHISHISLEUYRVALYASERTYRVALAAN 127
DB 1567 TTGGAAGGACCAAGACCTCCCTTTAGACATCACTGTAGTCACTTAACGTAAAT 1626
QY 128 PROGLYGLUVALTHRARGLEUTHRASPARGLIYTYRSEHISSEYSCYSLLESERGIN 147
DB 1627 CCTGAGAGGTACAAAGCTGACCTGACCTGCTGCTACATTTCTTGCTGATCAGTCA 1686
QY 148 HISYASPPHEPHILESERIYTYRSEANGINLYASNPROIHCYSVALSERLEU 167
DB 1687 CACTGTGACTCTTTTATAGTATAGTATAGTACCAAGAAATCCACAGTGTGCTCCTT 1746
QY 168 TYRYSLEUSERSERPROGLIUSASPSERPROTHCYSLYTHLYGILUHERPALATNR 187
DB 1747 TACAAGCTATCAAGTCTGAAGATGACCAACTTCGCAAAACAAAGAAATTTGGGCGACC 1806
QY 188 ILELEUASPSERIALGIYPROLEUPROASPIYRTHPRPROGLIUILEPHESERPHGIN 207
DB 1807 ATTTGGATTCAGCAGGTCCTCTTCTGACATATCTCCCGCAAAATTTCTTTTGA 1866
QY 208 SERTHRGILYPHETHRLEUITYGILEUITYLYSPROHISASPLEUNGIPROGIY 227
DB 1867 AGTACTAGTGAATTAATGATGATGAGATGCTCTCAAGACCTCATGATCTACAGCTGGA 1926
QY 228 LYSLYSTYRPROTHRVALLEUPHEIILETYRGIVGLYPROGINVALGILEUVALJASNA 247
DB 1927 AAGAAATTCCTACGTGCTGTTCATATATAGTGTCTCCAGGTGCAAGTGTGGAATAT 1986
QY 248 ARGPHELYSGIYVALYTYRPHENRGLUASNPHEIILESERLEUGIYTYRVALVAL 267


```
Db 1330 CACGAGAGAAATTGAGTTATTTTGGCTCTGATGCAAAACAGTTCCGTCATTTA 1389
QY 98 ----- 98
Db 1390 TACAAAATTACATCTATTTTAAAGAAACAAATATAACATCCAGTGTGGCTGCT 1449
QY 98 ----- 98
Db 1450 GCTCCAGTATTTCAAGTGTCTTCAAGAAGAGATGACATTCAGTGGTGAATGG 1509
QY 99 -----ValAspGluValArgLeuValTyr 107
Db 1510 GAAGTCTTGGCCGCGCATGATCTAATATCCAGTTGATGAAGTGAAGGCTGTATAT 1569
QY 108 PheGluGlyThrLeuAspSerProLeuGluHisIleuTyrValIleSerTyrValAsn 127
Db 1570 TTGGAAGGACCAAAAGACTCCCTTTAGAGCATCAGTCGTAGTCAGTTACGTAAT 1629
QY 128 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysSerIleSerGln 147
Db 1630 CCTGGAGAGGTGACAAAGGCTGACTGACCGTGGCTACTACATTTCTGCTGACGCTAG 1689
QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db 1690 CACTGTGACTTCTTATATAGTAAAGTATAGTAAACGAAAGAAATCCACACTGTGTCTCCTT 1749
QY 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrAlaThr 187
Db 1750 TACAAGCTATCAAGCTCGAAGCATGACCAACTTCGCAAAACAAAGAAATTTGGCCAC 1809
QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db 1810 ATTTGGATTGACGAGGTCCTCTCTGACTACTCTCTCCGAAATTTCTCTTTGAA 1869
QY 208 SerThrThrArgLysPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db 1870 AGTACTACTGAGTATTACATTTGATGAGATGCTCTCAAGACCTCATGATCTACAGCTGGA 1929
QY 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
Db 1930 AAGAATATCTTACTACGTGCTGTTCAATATGATGTGTCTTCAGAGTGCAGTTGTGAATAT 1989
QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db 1990 CGGTTTAAAGAGTCAAGATTTCCGCTTGAAATACCTTACGCTCTCTAGTTATGTGTT 2049
QY 268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
Db 2050 GTAGGATAGACAAACAGGGGATCCGTGACCCAGGGGCTTAAATTTGAAAGGCCCTTTAAA 2109
QY 288 TyrLysMetGlyGlnIleGlnIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db 2110 TATATAAATGGGTCAAATAGAAATTAACATCAGGTGAGAGGCTCCAAATATCTTACTCT 2169
QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyLysTyr 327
Db 2170 CGATATGATTTTCACTTACATGATCTGTGGGCAATCCAGGTGTGTCATGAGAGATAC 2229
QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db 2230 CTCTCCCTGATGCGATTAATGACAGGTGATATCTTACAGGGTGTGCTATGTGGGGCC 2289
QY 348 ProValThrLeuTyrIlePheTyrAspThrArgLysTyrThr 360
Db 2290 CCAATCACTCTGTGATCTTCTATATACAGATACAGATACAG 2328
RESULT 5
ABK83322
ID ABK83322 standard; cDNA; 2671 BP.
XX AC ABK83322;
XX
```

```
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPPIV related serine protease DPP-1.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200239134-A2X.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR ) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI, 2002-444176/47.
XX
DR P-PsDB; ABG61591.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
XX
XX Claim 1; Page 53-54; 113pp; English.
XX
PS
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPPR proteins.
XX
SQ Sequence 2671 BP, 805 A; 524 C; 594 G; 748 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 1,21e-204 Length: 2671
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Gaps: 173
DB: 24
US-10-070-464-7 (1-360) x ABK83322 (1-2671)
QY 1 GlnGluAspAlaArgSerIleGlyValAlaThrPheValLeuGlnGluLysPheAspArg 20
Db 737 GAGGAGAGTGGCAATGCAAGTGGAGTCCCTTGTCTTCCAGAAAGAAATTTATATGA 796
QY 21 TyrSerGlyTyrTyrProCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
Db 797 TATCTGCTATTTGCTGTGTCGCAAAAGCTGAACAACTCCAGTGTGTAATTTCT 856
QY 41 ArgIleLeuTyrGlnGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
```

Db 857 AGAATTCATATGAGAAAATGATGATCTGAGGTGAGAAATATTATTCATGTTACATCCCT 916
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProIySThrGlyThrAlaAsnPro 80
 Db 917 ATGTGGAACAAGAGGAGGAGATTCATTCGGTATCCTAAACAGGATACGCAAACTCT 976
 QY 81 LysValIlePheIlePheMetSerGluIleMetIleAspAlaGluGlyValIle----- 98
 Db 977 AAAGTCACTTTTAAAGATGTGAGAAATATGATGATGCTGAGAGAGATCATGATGTC 1036
 QY 98 ----- 98
 Db 1037 ATAGATAAGAACTAATTCACCTTTTGAGATTCATTTTGAGAGAGTTGAATATATTGCC 1096
 QY 98 ----- 98
 Db 1097 AGAGCTGATGAGCTCCTGAGAGAAATATGCTTGGTCATCTCACTAGATGCTCCAG 1156
 QY 98 ----- 98
 Db 1157 ACTGCTGAGATAGTGTGATCTCACCTGAATATTATTCAGAGAGATGATGTT 1216
 QY 98 ----- 98
 Db 1217 ATGAAAAGGAGAGACTCATGAGTCACTGCTGATTCGTGAGCCACTAATTATCTAT 1276
 QY 98 ----- 98
 Db 1277 GAAGAAACAACAGACATCTGATAAATATCCATGACATCTTTCANGTTCCTCCCAAGT 1336
 QY 98 ----- 98
 Db 1337 CACGAGAGAGAAATTGAGTTATTTTTCCTGATGCAAAACAGGTTTCGTCATTTA 1396
 QY 98 ----- 98
 Db 1397 TACAAAATTACATCTATTTTAAAGAAAGCAAAATATTAACGATCCAGTGGTGGCTGCT 1456
 QY 98 ----- 98
 Db 1457 GCTCAGAGTATTTCAAGTGTCTATCAAAAGAGATAGCAATTAACAGTGTGATGAG 1516
 QY 99 -----ValAspGluValAlaGluValIleTyr 107
 Db 1517 GAAGTCTTGCCGCGCATGATCTAATATCCAAAGTTATGAGAGCAGAGCTGTATAT 1576
 QY 108 PheGluGlyThrIleAspSerProLeuGluIleHisIleuTyrValIleSerTyrValAsn 127
 Db 1577 TTTGAGGACCAAGACTCCCTTTAGAGCATACCTGTCAGTCACTTACGTTAAT 1636
 QY 128 ProGluGluValIleThrArgLeuThrAspArgGlyTyrSerHisSerCysAsiIleSerGln 147
 Db 1637 CCTGAGAGGCTGACAGAGCTGACGCTGCTACACACTTCCTGTCATCAGTCAG 1696
 QY 148 HisCysAspPhePheIleSerIleTyrSerAsnGlnIleAsnProHisCysValIleSerLeu 167
 Db 1697 CACGTACTCTTTATATAGATATATAGTATACAGAGAAATCCACCTGTGTCTCTT 1756
 QY 168 TyrIleuSerSerProGluAspAspProThrCysIleThrIleGluPheTyrAlaThr 187
 Db 1757 TACAGCTATCAATCTCTGAAAGATGACCCAACTTGCAAAACAAAGAAATTTGGGCCACC 1816
 QY 188 IleIleuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
 Db 1817 ATTTTGATTCAGAGGCTCTCTCTGACATATCTCTCCAGAAATTTTCTCTTTGAA 1876
 QY 208 SerThrThrGlyPheThrIleuTyrGlyMetIleuTyrIleProHisAspLeuGlnProGly 227
 Db 1877 AGTACTACTGATTTACATATGATGAGATGCTCTCAACCTCAAGATCTACACCTGGA 1936
 QY 228 LysIleTyrProThrValIleuPheIleTyrGlyGlyProGluValIleGluValIleAsn 247
 Db 1937 AAGAAATATCTACTGTGCTGTTCATATATGCTGCTCTCAGGTCAGGTTGGTGAATAT 1996

QY 248 ArgPheIleGlyValIleTyrPheArgLeuAsnThrIleuIleSerLeuGlyTyrValIle 267
 Db 1997 CGATTTAAAGAGCAAGTATTTCCGCTTAATACCTAGCTCTCTAGGTTATGTGTT 2056
 QY 268 ValValIleAspAsnArgIleSerCysHisArgGlyIleuIlePheGluGlyValIlePheIle 287
 Db 2057 GTAGTGTATACCAACAGAGGAGCTCTGACCGAGGCTTAAATTTGAAGGGCCCTTAAA 2116
 QY 288 TyrIleuMetGlyIleGluIleAspAspGlnValIleGluIleuGlnTyrIleuIleSer 307
 Db 2117 TATTAATGGGTCAATAGAAATGACGATCAGGTGAGAGACTCCATATCTAGCTTCT 2176
 QY 308 ArgTyrAspPheIleAspLeuAspArgValIleGlyIleHisGlyTyrPheTyrGlyIleTyr 327
 Db 2177 CGATATGATTTTCATTTGACTTATGATCGTGGGACATCCAGCGCTGCTCTATGAGAGATAC 2236
 QY 328 LeuSerIleuMetAlaLeuMetGlnArgSerAspIlePheArgValIleAlaGlyValIle 347
 Db 2237 CTCTCCCTGATGGGATTAATGACAGAGTCAAGATATCTTCAAGGTTGCTATTGCTGGGCC 2296
 QY 348 ProValThrIleuThrIlePheTyrAspThrGlyTyrThr 360
 Db 2297 CCACTCACTCTGTGATCTTCTATGATACAGATACAG 2335
 RESULT 6
 ID ABRK12892 standard; cDNA; 3106 BP.
 AC ABRK12892;
 XX
 AC ABRK12892;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Human protease PRTS-9 cDNA sequence.
 DE
 XX Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
 KM cardiovascular; atherosclerosis; autoimmune deficiency syndrome; AIDS;
 KM inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
 KM cell proliferative disorder; developmental disorder; epilepsy;
 KM Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
 KM reproductive disorder; endometriosis; ss.
 XX
 OS Homo sapiens.
 XX
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 203..2851
 FT /*tag= a
 FT /product= "Human protease PRTS-9"
 FT
 XX
 PN MO200198468-A2.
 XX
 XX
 PD 27-DEC-2001.
 XX
 PD 13-JUN-2001; 2001WO-US19178.
 XX
 PR 16-JUN-2000; 2000US-212336P.
 XX
 PR 22-JUN-2000; 2000US-213955P.
 XX
 PR 29-JUN-2000; 2000US-215396P.
 XX
 PR 07-JUL-2000; 2000US-216821P.
 XX
 PR 14-JUL-2000; 2000US-218946P.
 XX
 FA (INCY-) INCYTE GENOMICS INC.
 XX
 XX Yue H, Elliott VS, Gandhi AR, Lai P, Au-young J, Tribouley CW;
 PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
 PI Wajia NK, Yag MG, Lu DM, Patterson C, Tang YT, Walsh RT;
 PI Azimzai Y, Li Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
 PI Kallick DA;
 XX
 DR MPI; 2002-090437/12.
 DR P-PSDB; AAU74749.
 XX
 FT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful

in the diagnosis, treatment and prevention of gastrointestinal (e.g. gastritis), cardiovascular (e.g. atherosclerosis) and cell proliferative (e.g. cancer) disorders -

Claim 5; Page 166-167; 177pp; English.

The present invention relates to twenty one new human proteases, referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and polypeptides of the invention are useful in the diagnosis, treatment and prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and myocardial infarction, autoimmune/inflammatory e.g. acquired immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell immunodeficiency e.g. cancer, developmental e.g. Duchenne and Becker muscular dystrophy, epithelial e.g. dermatitis, neurological e.g. epilepsy and Alzheimer's disease and reproductive e.g. infertility and endometriosis disorders. Numerous other examples of each disorder are given in the specification. The present nucleic acid sequence encodes the human protease PRTS-9 protein of the invention.

Sequence 3106 BP, 928 A, 633 C, 704 G, 841 T, 0 other;

Alignment Scores:
Pred. No.: 1,52e-204 Length: 3106
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 24 Gaps: 1

US-10-070-464-7 (1-360) x ABK12892 (1-3106)

```

OY 1  GUGUAGSPLAAGSERSLAGLYVALAATHRPHVALLEUINGLUGLPHASPARG 20
DB 932 GAAGAGAGTGCAGATCGAGAGTCGCTACCTTGTCTCCAGAAATTGATGGA 991
OY 21  TYSERGIYTYRTPRCPYSPROLYSALAGLUTHRPHROSERGIYLYSILEU 40
DB 992 TATTCGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1051
OY 41  ARGILEUTYRGUGLUGLUSNAPSLUSERGIYVALGULELLEHISVALTHSERP 60
DB 1052 AGAATTCTATATGAAGAAATGATGATGCTGAGGGAATTTATTCATGTTACATCCCT 1111
OY 61  METLEUGLUTHRARXARXALASPSEPHARGTYRPROLYSTHGLYTHRALASNP 80
DB 1112 ATGTTGGAACAGAGGAGGAGATTCCTGTTATCCTAACACGTRACGCAATCCT 1171
OY 81  LYSVALTHRPHLYSMETSERGIULEMETILLEASPLAGLUGLYARGILEU 98
DB 1172 AAAGTCATTTTAAGATGTCAGAAATATGATGATGCTGGAAGAGATCATGATGTC 1231
OY 98  ----- 98
DB 1232 ATAGATAAGAACTAATTCAACTTTTGAGATTCTATTGGAAGATTGAATATATGCC 1291
OY 98  ----- 98
DB 1292 AGAGCTGATGAGCTCTCTGAGGAAATATGCTTGCTCATCTACTAGATGCTCCAG 1351
OY 98  ----- 98
DB 1352 ACTGCGCTACAGATAGTGTGATCTCACCTGAATATTATTATCCAGTAGAAGATGATGTT 1411
OY 98  ----- 98
DB 1412 ATGGAAGGACAGAGACTCATTTGAGTCACTGCTGATCTGTGACGCCACTAATTATCTAT 1471
OY 98  ----- 98
DB 1472 GAAGAAACAGACATCTGATTAATATTCATGATCATCTTTTATCCCAAGT 1531
OY 98  ----- 98

```

```

DB 1532 CACGAAGAGAAATTGAGTTATTTTGGCTCTGAAATGCAAAAGAGTTCCGTCATTTA 1591
OY 98  ----- 98
DB 1592 TACAAATATCATCTATTTTAAAGAAAGCAATATATAAGATCCAGTGTGGCTGCT 1651
OY 98  ----- 98
DB 1652 GCTCCAGATGATTTCAAGTGTCTTATCAAGAGAGATAGCAATTACCAGTGTGAATGG 1711
OY 99  ----- 1711
DB 1712 GAAGTCTTGCCCGCAGATGATCTAATATCCAAAGTTATGAAAGTCAGAGGCTGTATAT 1771
OY 108 PHEGLUGLYTHRYSASPSEPROLEUGLHHISHILEUTYRVALSERTYRVALASN 127
DB 1772 TTGAGGACCAAGAACTCCCTTTAGAGCATCAGCTGACGTACGTACGTACGTACGTAAAT 1831
OY 128 PROGLYGLVALTHRAGLEUTHRASPARGLYTYRSEHISSERCYSYSILLESERGIN 147
DB 1832 CCTGAGAGGTGACAGAGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 1891
OY 148 HISCYASAPHEPHEILLESERYTYRSEASNGINLYSASNPROHISCYVALSERLEU 167
DB 1892 CACTGTGACTCTTTATATAGATATAGTATACAGAGAAATCCACACTGTGTCTT 1951
OY 168 TYRILYSSEUSERPROGLUSPASPPTHRCYSILSYTHIRLYSGLPHETRALTHR 187
DB 1952 TACAAGCTATCAAGCTCGAAGATGACCAACCTTGCAAAACAAAGAAATTTTGGCCACAC 2011
OY 188 ILELEUSPSERALAGLYPROLEUPROASPYRTHRPROPROGLULLEPHESEPHGLU 207
DB 2012 ATTTTGGATCAGAGGCTCTCTCTGACTATACCTCCAGAAATTTTCTTTTGA 2071
OY 208 SERTHRTHRGYPHEHLEUTHIRLYGMELEUTHIRLYSPROHISASPSELEUGLNP 227
DB 2072 AGTACTGATGATTTACTTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2131
OY 228 LYSILSYRPROTHVALLEUPHEILETYRGLYGLYPROGLINVALGLEUVALASN 247
DB 2132 AAGAAATATCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2191
OY 248 ARGPHLYSGLYVALSYRPHARXLEUASNTHIRLEULASERLEUGLYTYRVAL 267
DB 2192 CGATTTAAGAGTCAAGATATTTCCGCTTGAAATCCCTAGCCTCTTAGGTATAGTGCTT 2251
OY 268 VALVALLEASPASNARGLYSERCYSHISARGLYLEULYSPEGLIYVALAPHELYS 287
DB 2252 GTAGGATATGACCAAGGAGATCCTGTCACCGAGGCGTTAATTTGAAGGCGCTTTAA 2311
OY 288 TYRILYSMETGLYGINILEGLULLEASPAPGLINVALGULYLEUGINTYRLEULASER 307
DB 2312 TATAAAATGGGTCAAAATGAAATGACGATGAGTGGAAGACATCCATATCTAGCTTCT 2371
OY 308 ARGTYRASPHEILEASPLEUSPARVALGILYILEHISGLYTPSERTYRGLYGLYTYR 327
DB 2372 CGATATGATTTTCAATGACTTATGATCGTGTGGCATCCACGCTGTCTTATGAGAGATAC 2431
OY 328 LEUSERLEUWEALALEUMECGLNARGSERASPILEPHARGVALAILEALAGLYALA 347
DB 2432 CTCTCCCTGATGGCATTAATGACAGAGGTCAAGATATCTTCAAGGATGTGATGTGGGGCC 2491
OY 348 PROVALTHIRLEUTHIRPHETTYRASPETHIRGLYTYRTHR 360
DB 2492 CCACTCACTGTGATCTTCTATGATATACAGAAATACAG 2530

```

RESULT 7

ID AAC85694 standard; cDNA; 3120 BP.

AC AAC85694;
XX

DT 29-JUN-2001 (first entry)

XX Nucleotide sequence of human DPP8.

DE Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
 KM dipeptidyl peptidase; DPP8; T cell; cleavage; diarrhoea;
 KM growth hormone deficiency; glucose level; mucosal regeneration;
 KM non-insulin dependent diabetes mellitus; glucose intolerance;
 KM immunosuppression; ss.

XX Homo sapiens.

OS

XX Key Location/Qualifiers

FT CDS 214..2862

FT /tag= a

FT /product= "Human DPP8"

XX WO200119866-A1.

XX 22-MAR-2001.

XX 11-SEP-2000; 2000WO-AU01085.

XX 10-SEP-1999; 99AU-0002762.

PR 18-FEB-2000; 2000AU-0005709.

XX (UNSY) UNIV SYDNEY.

PA Abbott CA, Gorell MD;

PI WPI: 2001-281520/29.

XX P-PSDB; MAB47187.

DR

XX New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
 PT substrates, identifying inhibitors of DPP8 catalytic activity which
 PT have therapeutic uses, and for detecting activated T cells

XX Claim 16; Fig 2; 78pp; English.

PS

XX This sequence encodes human dipeptidyl aminopeptidase (DPP8). DPP8
 CC has substrate specificity for H-Ala-Pro-pNA, H-Gly-Pro-pNA and
 CC H-Arg-Pro-pNA. Therefore, it is a prolyl oligopeptidase and a
 CC dipeptidyl peptidase, because it is capable of hydrolysing the
 CC peptide bond C-terminal to Pro in each of these compounds. DPP8
 CC is homologous with human DPP8. DPP8 is useful for cleaving a
 CC substrate, and for detecting an activated T cell which involves
 CC measuring the level of DPP8 gene expression in a T cell. The level
 CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
 CC in the cell. It is also useful for identifying a molecule capable
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules
 CC identified as inhibiting DPP8 catalytic activity may be useful for
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
 CC in non-insulin dependent diabetes mellitus and other disorders
 CC involving glucose intolerance, enhancing mucosal regeneration and
 CC as immunosuppressants.

XX

SQ Sequence 3120 BP, 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:

Pred. No.:	1,536-204	Length:	3120
Score:	1836.50	Matches:	360
Percent Similarity:	67.54%	Conservative:	0
Best Local Similarity:	67.54%	Mismatches:	0
Query Match:	95.01%	Indels:	173
DB:	22	Gaps:	1

US-10-070-464-7 (1-360) x AAC85694 (1-3120)

QY 1 GIUGLUSPALAAGSERIAGLYVALAAlaThrPheValLeuGlnGluIlePheASParg 20

DB 943 GAAGAATAATGACCATGAGTGGAGTGGCTACCTTTGTTCTCCCAAGAGATTGATAGA 1002

QY 21 TysSerGlyTyrTrpTyrCysProIysAlaGluThrThrProSerGlyGlyLeu 40

DB 1003 TATTCGCTATATGATGATGCTCAAAAGCTGAACCACTCCAGTGGTGAATATTTT 1062

QY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleHisValThrSerPro 60

DB 1063 AGATTCTATATGAAGAAATGATGATCTGAGTGGAAATATTATCATGTTACATCCCT 1122

QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProIysThrGlyThrAlaAsnPro 80

DB 1123 ATGTGGAAACAAAGAGAGGAGATTCATTCCTGATCTCAAAACAGGTACGCAATCT 1182

QY 81 LysValThrPheIysMetSerGluIleMetIleAspAlaGluIleLeu----- 98

DB 1183 AAAGTCACTTTAAGATGTCAGAAATATGATGATGCTGGAAGAGATCATGATGTC 1242

QY 98 ----- 98

DB 1243 ATAGATAAGACATTAATTCACCTTTGAGATTCATTTTGAAGAGATTGAATATATTC 1302

QY 98 ----- 98

DB 1303 AGAGCTGATGAGACTCCTGAGGAAATATGCTGTCATCTAGATGCTGCCAG 1362

QY 98 ----- 98

DB 1363 ACTGCCTAGAGATGATGTTGATCTCACTGAATTATTTATCCAGTAAAGATGATGTT 1422

QY 98 ----- 98

DB 1423 ATGGAAGAGCAGAGACTCATGAGTCACTGCTGATTCGTGAGCCCATTAATTCAT 1482

QY 98 ----- 98

DB 1483 GAAGAAACAACAGACATCTGGATTAATATCCATGACATCTTTCATGTTTCCCAAGT 1542

QY 98 ----- 98

DB 1543 CACGAAGAGAAATGAGTTATTTTGTGCTGTAATGCAAAACAGTTTCGTCATTTA 1602

QY 98 ----- 98

DB 1603 TACAATAATTCATCTATTTTAAAGAAAGCAAAATATAACATCCAGTGGGTGCT 1662

QY 98 ----- 98

DB 1663 GCTCAAGATGATTTCAAGTGTCTATCAAGAGAGATAGCAATTACCAAGTGTGATGG 1722

QY 99 -----ValaspGluValArgArgLeuValTyr 107

DB 1723 GAAGTTCTTGCCGCGCATGATCTAATATCCAAAGTTGATGAAGTCAGAGCTGTATAT 1782

QY 108 PheGluGlyThrIysAspSerProLeuGluHisHisLeuTyrValIleSerTyrValAsn 127

DB 1783 TTTGAAGGACCAAGAACTCCCTTTAGAGATCACTGTCGTAAGTACGTAATAT 1842

QY 128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGln 147

DB 1843 CCTGGAAGGTGACAAAGCTGACGACCGTGTACTACATCTCTTCGATCAGTACG 1902

QY 148 HisCysAspPhePheIleSerIysTyrSerAsnGlnIysAsnProHisCysValSerLeu 167

DB 1903 CACTGTACTCTTTAATAGATAGTATAGTACCAAGAGATCCACACTGTGTCTCTT 1962

QY 168 TyrIleLeuSerSerProGluAspAspProThrCysIysThrIysGluIlePheSerPheGlu 187

DB 1963 TACAGCTATCAAGTCTCGAAGATGACCAACTTGCAAAACAAAGAAATTTGGGCCACC 2022

QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207

DB 2023 ATTTTGATTCAGAGGTCTCTTCTGACTATCTCTCAGAAATTTTCTTTTGA 2082

QY 208 SerThrThrGlyPheThrIleLeuTyrGlyMetLeuTyrIysProHisAspLeuGlnProGly 227

Db 2083 AGTACTGATGATTACATTTGATGGATGCTCTACAAAGCCTCATGATCTACAGCCTGGA 2142
QY 228 LysLysIYrProThValleuPheIleTyrgIyglYProGlnValGlnleuValasn 247
Db 2143 AAGAAATATCTTACTGCTGCTGCTATATATGATGCTCTCAGAGTCAAGTGTGATATAT 2202
QY 248 ArgPheLysGLValIleTyrrPheArgLeuAenThrLeuAlaSerLeuGLYTyrrValVal 267
Db 2203 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTTACCTCTCTAGTTATGTGTT 2262
QY 268 ValValIleAspAsnArgIYSerCysHisArgIYLeuLysPheGlnGLYAlaPheLys 287
Db 2263 GTAGGATAGACAAACAGGGGATCTGTGACCCAGGGGCTTAATTTGAAGGCCCTTTAA 2322
QY 288 TyrrLysMetGLYglnIleGlnIleAspAsnGlnValGlnGLYLeuGlnTyrrLeuAlaSer 307
Db 2323 TATATAAATGGGTCAATATGAATATGACATCAGTGAAGGACTCCAAATATCTACCTCT 2382
QY 308 ArgTyrrAspPheIleAspLeuAspArgValGlyIleHisGLYTrpSerTyrgIyglYTyrr 327
Db 2383 CGATATGATTTGATTTGATTTGATGATGCTGTGGGATCCAGGCTGTGCTTATGAGAGATAC 2442
QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGLYAla 347
Db 2443 CTCTCCCTGATGGCATTAATGACAGAGTCCAGATATCTTCAGGGGTGCTATTGCTGGGCC 2502
QY 348 ProValThrLeuThrIlePheTyrrAspThrArgIYTyrrThr 360
Db 2503 CCAGTCACCTGTGTGATCTTCTATGATACAGGATACAGC 2541
RESULT 8
AAD38956
ID AAD38956 standard; cDNA, 3120 BP.
AC AAD38956;
AC AAD38956;
DT 23-SEP-2002 (first entry)
DE Human dipeptidyl peptidase 8 (DPP8) cDNA.
XX Human, dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW antiviral; enzyme; gene; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 214..2862
FT CDS /tag= a
FT /product= "Human DPP8 protein"
XX
FN WO200234900-A1.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-AU01388.
XX
PR 27-OCT-2000; 2000AU-0001078.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Abbott CA, Gorrell MD;
XX
DR WPI: 2002-45464/48.
DR P-PSDB; AAB24170.
XX
PT New dipeptidyl peptidase (DPP) peptides, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection -
XX
PS Example; Fig 1; 91pp; English.

XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC polymucleotides encoding such proteins. The DPP peptides are useful for
CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC rejection and HIV (human immuno deficiency virus) infection. The present
CC sequence is human DPP8 cDNA.
SQ Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;
Alignment Scores:
Pred. No.: 1.53e-204 Length: 3120
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 173
Query Match: 95.01% Indels: 1
DB: 24 Gaps: 1
US-10-070-464-7 (1-360) x AAD38956 (1-3120)
QY 1 GlnGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluPheAspArg 20
Db 943 GAAGAAAGTGGCAATCAGCTGAGATCGCTACCTTGTCTCCAGAAAGAAATTGATGA 1002
QY 21 TyrrSerGlyTyrrTrpCysProLysAlaGlnThrThrProSerGlyGLYIleLeu 40
Db 1003 TATCTGGCTATTGGTGTGTGCTCCAAAGCTAAACACTCCAGTGTGTAAATTCCT 1062
QY 41 ArgIleLeuTyrgIygluLysAsnAspGlnSerGlnValGlnIleHisValThrSerPro 60
Db 1063 AGAATTCATATGAAGAAATGATGAACTCAGGTGGAATTAATTCATGTTACATCCCT 1122
QY 61 MetLeuGlnThrArgArgAlaAspSerPheArgTyrrProLysThrAlaAsnPro 80
Db 1123 ATGTTGGAAACAGAGGCGAGATTCATTCCTTAATCTTAAACAGTACGAAATCCT 1182
QY 81 LysValThrPheLysMetSerGlnIleMetIleAspAlaGlnGLYArgIleIle----- 98
Db 1183 AAAGTCATTTTAAAGATGTCGAAATATGATGATGCTGTAAGAGAGATCATTAATGTC 1242
QY 98 ----- 98
Db 1243 ATGATTAAGAACTAATTCACCTTTGAGATTCATTTGAAGAGATTGAATATATGCGC 1302
QY 98 ----- 98
Db 1303 AGAGCTGATGACTCTGAGAGAAATATGCTGTCCATCTATAGATGCTCCAG 1362
QY 98 ----- 98
Db 1363 ACTGGCTTACGATATGATGATCTCAGCTGAATTTATTTATCCAGTGAAGATATGTT 1422
QY 98 ----- 98
Db 1423 ATGGAAGGAGAGACTCATTTAGTCAAGTGCCTGATCTGTGAGCGCACTAATATCTAT 1482
QY 98 ----- 98
Db 1483 GAAGAAACACAGACATCTGATTAATATCCATGACATCTTCAATGTTTTTCCCAAGT 1542
QY 98 ----- 98
Db 1543 CACGAAAGAAATTGAGTTTATTTTCCCTGTAATGCAAAAACAGTTTCCGTCATTTA 1602
QY 98 ----- 98
Db 1603 TACAAATATCATCTATTTTAAAGAAAGCAATATTAACGATCCAGTGTGCTGCT 1662
QY 98 ----- 98
Db 1663 GCTCCAAAGTATTTCAAGTGTCTTATCAAAAGAGAGATGCAATTAACAGTGTGAATG 1722
QY 99 -----ValAspGlnValArgArgLeuValTyrr 107

```

Db      1723 GAAGTCTTGCGCCGCGATGATCTAAATCCAAAGTTGAAGAGCTGAGATTTAT 1782
Qy      108 pHeGluGlyThrLysAspSerProLeuGluHisLysLeuValLysTyrValAsn 127
Db      1783 TTTGAAGGACCAAGACCTCCCTTTAGAGCATCACTGAGCTAGTCAAGTTACGTAAT 1842
Qy      128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyIleSerGln 147
Db      1843 CCGGAGAGAGTGCACAAAGGCTGACTGACCGGGCTACTCATCTTGCTGCACACTCAG 1902
Qy      148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAspProHisCysValSerLeu 167
Db      1903 CACTGTGACTCTCTTTTAAGTAAGTAATGATACCAAGATCACTGAGTGTCTCTT 1962
Qy      168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 187
Db      1963 TACAAGCTATCAAGTCTCGAAGATGACCCCAACTGCAAAACAAAGAAATTTGGGCAC 2022
Qy      188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db      2023 ATTTGGATTTCAGCAGAGTCTCTCTCTGACTATACCTCCCAAAATTTTCTTTTGA 2082
Qy      208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db      2083 AGTACTACTGATTTTCATTTGATGAGATGCTCTACAGAGCTCATGATCTACAGCTTGA 2142
Qy      228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 247
Db      2143 AAGAAATATCTACTGCTGCTGCATATATAGTGCTCTCAGAGTGCAAGTTGGATTAAT 2202
Qy      248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db      2203 CGGTATTAAGAGATCAAGATATTCCTGCTGAAATCCCTGCTAGGATATGCGT 2262
Qy      268 ValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
Db      2263 GATGATATGACACACAGGGATCTCTTACCGAGGCTTAATTTAAAGCCCTTTAA 2322
Qy      288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGlnGlyLeuGlnTyrLeuAlaSer 307
Db      2323 TTTAAATGGGTCAATATGAATTTGACGATCAGGTGGAAGACTCCATATCTAGCTTCT 2382
Qy      308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyTyr 327
Db      2383 CCAATATGATTTCACTTACTTAAATCGTGGGATCCAGCGCTGCTTATGAGGATAC 2442
Qy      328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db      2443 CTCTCCCTATGATTAATGACAGAGTACAGATATCTTACAGGTTCTATTTGCTGAG 2502
Qy      348 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
Db      2503 CCAGTCACTCTGTGATCTTCTATGATACAGGATACAG 2541

RESULT 9
AAH99934
ID      AAH99934 standard; cDNA; 3143 BP.
AC      AAH99934;
XX
XX
DT      12-APR-2002 (first entry)
XX
XX
DE      cDNA encoding 21953 human prolyl oligopeptidase.
XX
XX      21953 prolyl oligopeptidase; human; proline; endopeptidase;
KW      cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW      neuronal disorder; vascular disorder; prostate disease; cytostatic;
KW      antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
KW      diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW      Grave's disease; neuronal disorder; demyelinating disease; ss.
XX
OS      Homo sapiens.

```

```

XX      Key Location/Qualifiers
FH      CDS 229..2877
FT      /tag= a
FT      /product= "21953 prolyl oligopeptidase"
FT      /note= "This region is specifically claimed in
FT      0 claim 2"
PN      WO200179473-A2.
XX
XX      25-OCT-2001.
PD
XX
XX      11-APR-2001; 2001WO-US40483.
XX
XX      18-APR-2000; 2000US-197508P.
XX
XX      (MILL-) MILLENNIUM PHARM INC.
PA
XX
PI      Meyers RA, Williamson M;
XX
XX      WPI; 2002-034353/04.
XX      P-PSDB; AAG78415.
XX
XX      New polypeptides 21953, member of human prolyl oligopeptidase family,
XX      useful as diagnostic targets and therapeutic agents for controlling
XX      cancer, lymphoma and leukemia
XX
XX      Claim 7; Page 100-102; 121p; English.
XX
XX      This invention relates to an isolated 21953 human prolyl
XX      oligopeptidase. Which is cyostatic, antidiabetic, antiarthritic,
XX      neuroprotective, antihypoid, dermatological, antipsoaritic,
XX      antiasthmatic, ophthalmological, antiinflammatory, nootropic,
XX      antiparinsonian, anticonvulsant, gynecological, vasotropic,
XX      antitraginal, cardiant, antiatherosclerotic, anorectic and
XX      metabolic in its action. Uses include gene therapy, expression or
XX      compund of 21953 protein modulator, it is useful for identifying a
XX      activity which binds to it and can be used in preventing, treating
XX      or detecting a cellular proliferative or differentiative disorder.
XX      The 21953 molecules can act as novel diagnostic targets and therapeutic
XX      agents for controlling disorders associated with the aberrant activity
XX      or degradation of peptide hormones e.g., disorders associated with cell
XX      differentiation and proliferation such as cancer, immune function,
XX      reproductive, neurological and cardiovascular function. The 21953
XX      molecules are thus useful for treating and preventing cellular
XX      proliferative and differentiative disorders, haematopoietic neoplastic
XX      disorders, immune disorders such as autoimmune diseases, diabetes
XX      mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
XX      neuronal disorders, demyelinating diseases, vascular disorders and
XX      metabolism or pain disorders. This sequence represents the cDNA
XX      encoding sequence of 21953 human prolyl oligopeptidase.

SQ      Sequence 3143 BP; 943 A; 644 C; 712 G; 844 T; 0 other;

Alignment Scores:
Pred. No.: 1.54e-204 Length: 3143
Score: 1636.50 Matches: 360
Percent Similarity: 67.54% Conservatve: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 24 Gaps: 1

US-10-070-464-7 (1-360) X AAH99934 (1-3143)
Qy      1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
Db      958 GAAGAAGAGCCAGATACCTGAGTGGCTACTTTGTTCTCCAAACAATAATTTGATACA 1017
Qy      21 TyrSerGlyTyrTyrTyrPyrCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
Db      1018 TATTCGCTATTTGGTGGTGTCCAAAGCTGAACAACATCCACAGTGTGTAAATTTCTT 1077
Qy      41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60

```



```

Db      1078 AGAATCTATATAGAAATGATGATCGAGGTGGAATTAATCATGTTACATCCCT 1137
Qy      61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
Db      1138 ATCTGGAAACAAAGAGGCGAGATTCATCCGTTATCTTAAACAGGTACAGCAAAATCCT 1197
Qy      81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyIleIle----- 98
Db      1198 AAAGTCACTTTTAAGATGCAAGAAATATGATGATGCTGAGGAGAGATCATGATGTC 1257
Qy      98 ----- 98
Db      1258 ATGATTAAGAACTAATTCACACCTTTGAGATTTCTATTGAAGAGTTGATATATTGCC 1317
Qy      98 ----- 98
Db      1318 AGAGCTGATGAGCTCCTGAGGAAATATAGCTTGTCATCTACTAGATCGCTCCAG 1377
Qy      98 ----- 98
Db      1378 ACTGCGCTGAGATAGTGTGATGTCACCTGAATTAATTATCCAGTGAAGATGATGTT 1437
Qy      98 ----- 98
Db      1438 ATGGAAGGAGAGACTCATTTAGTCACTGCGCTGATTCGTGAGAGCCACTAATTAATCTAT 1497
Qy      98 ----- 98
Db      1498 GAAGAAACAAAGACATCTGATTAATATCCATGATGATCTTCAATGTTTTCCCAAGT 1557
Qy      98 ----- 98
Db      1558 CACGAAGAGAAATTTGAGTTATTTTTCCTCTGAAATGCAAAACAGSTTCCGTCATTTA 1617
Qy      98 ----- 98
Db      1618 TACAAATATCATCTATTTTAAAGAAAGCAATATATAAGATCCAGTGGGCTGCTCCT 1677
Qy      98 ----- 98
Db      1678 GCTCCAAAGTATTCAAGTGTCTATCAAAAGAGATGACATTCACAGTGTGAATGG 1737
Qy      99 ----- ValAspGluValArgArgLeuValTyr 107
Db      1738 GAAGTCTTGCCCGCATGATCATATATCCAGTTATGAGTCAAGAGCGCTGATAT 1797
Qy      108 PheGluGlyThrLysAspSerProLeuGluHisIleuTyrValValSerTyrValAsn 127
Db      1798 TTGGAAGGACCAAAAGACTCCCTTTAGAGATCACCCTGTACGTACGTAAAT 1857
Qy      128 ProGlyValValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyIleSerGln 147
Db      1858 CCTGGAAGGCTGACCTGACTGACCGGCTACTCACTCACTTCTGCTGATGATGACG 1917
Qy      148 HisCysAspPhePheIleSerLysTyrSerAsnGluLysAsnProHisCysValSerLeu 167
Db      1918 CACTGATCTTCTTTATAGTAAGTATAGTAAACAGAAAGATCACACTGATGCTCCTT 1977
Qy      168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheIleThr 187
Db      1978 TACAAGCTATCAACTCTGAAAGATGACCAACTTGCAAAACAAAGGAATTTGGGCCACC 2037
Qy      188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db      2038 ATTTTGATTCAGAGAGTCTCTTCTGACTATACCTCTCCAGAAATTTTCTCTTTGAA 2097
Qy      208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db      2098 AGTACTACTGATTTATCATGTATAGGATGCTCTACAAAGCTCATGATCTACACCTGGA 2157
Qy      228 LysLysTyrProThrValIleuPheIleTyrGlyProGlnValGlnLeuValAsnAsn 247

```

```

Db      2158 AAGAATATCTACTGCTGCTCATATATGTTGTCCTCAGGTCGAGTTGTGAATAAAT 2217
Qy      248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db      2218 CGGTTTAAAGGATCAAGTAATTCGCTTGAATAACCTTAGCTCTAGGTTATGTGTT 2277
Qy      268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluValAlaPheLys 287
Db      2278 GTAGTATAGACACACGGGATCTCTCACCGAGGCTTAAATTTGAAGCGCTTTAA 2337
Qy      288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db      2338 TATTAATAGGCTCAATTAAGAAATTCAGATCAGGTGGAAGAGACTCCAATATCTAGCTTCT 2397
Qy      308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyrGlyTyr 327
Db      2398 CGATATGATTTCTATTAATTCATTAATCGTGGGATCCACGGCTGCTTATGAGGATAC 2457
Qy      328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db      2458 CTCTCCCTGATGCAATTAATGACAGGTCAAGATATCTTCAAGGTTGCTATGCTGGGCC 2517
Qy      348 ProValThrLeuTyrPheTyrAspThrGlyTyrThr 360
Db      2518 CCAGTCACTCTGTGATCTTCTATGATACAGATACACG 2556

RESULT 10
ABK83327 standard; cDNA; 4829 BP.
XX
AC ABK83327;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPP-1 splice variant #3.
XX
ID Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001MO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Juntien J;
XX
DR WPI; 2002-444178/47.
XX
DR P-PSDB; AB661596.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PS Disclosure; Page 65-66; 113pp; English.
XX
XX The present invention relates to the isolation of novel human serine
XX proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
XX proteases (DPP-IV). The dipeptidyl peptidase IV-related proteins (DPP-IV)
XX and nucleic acids encoding them are useful for treating infections
XX such as fungal, bacterial, protozoan and viral infections, particularly

```


PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 DR
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 XX Claim 8; SEQ ID 12963; 2537bp + CD ROM; English.
 XX
 XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 XX Sequence 2161 BP; 662 A; 429 C; 452 G; 618 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 1,38e-168 Length: 2161
 Score: 1528.00 Matches: 308
 Percent Similarity: 57.79% Conservative: 0
 Best Local Similarity: 57.79% Mismatches: 1
 Query Match: 79.05% Indels: 224
 DB: 22 Gaps: 2
 US-10-070-464-7 (1-360) x AAH15009 (1-2161)
 QY 1 GUGUASPAlaArgSerAlaGlyValAlaIlePheValLeuGlnGluGluPheAspArg 20
 DB 148 GAGGAAGATGCCAGATCAGCTGAGTCGCTACCTTGTCTCCAGAAATAATTGTGATAA 207
 QY 21 TysSerGlyTyrTyrTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
 DB 208 TATTCGCTATTGGTGGTGTCTCCAAAGCTGAAACCACTCCAGTGTGTAAATAATTCCT 267
 QY 41 ArgIleLeuTyrGlnGluLysAspGluSerGluValGlnIleIleHisValThrSerPro 60
 DB 268 ACAATTCATATGAAAGAAATGATGATCTGAGGTGGAATATTATTCATCATCCCT 327
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAspPro 80

DB 328 ATGTTGAAACAAGAGGGCAGATTCAATCCGTTATCTTAAACAAGGTACAGCAATCT 387
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
 DB 388 AAAGTCATCTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGATGATCATATGATGC 447
 QY 98 ----- 98
 DB 448 ATAGATMAAGAACTAATTCACCTTTTGATTTCTAATTTGAAGAGCTGAATATATATGCC 507
 QY 98 ----- 98
 DB 508 AAGAGCTGATGAGACTCTCTGAGGGAAATATGCTGTGTCATCTCTACTAGATGCTCCAG 567
 QY 98 ----- 98
 DB 568 ACTCGCTACAGATAGTGTGTGATCTCACCTGAATATTTATCCAGTAGAAGATGATGTT 627
 QY 98 ----- 98
 DB 628 ATGAAAGGACAGACTCTATTGATGATGCTGCTGATTCGTGACGCCACTAATATCTAT 687
 QY 98 ----- 98
 DB 688 GAAAGAAACAACAGACATCTGATMAAATATCATGACATCTTTCATGTTTCCCAAGT 747
 QY 98 ----- 98
 DB 748 CACGAGAGAGAAATGAGTTTATTTTGGCTGTGATGCAAAACAGTTTCGTCATTTA 807
 QY 98 ----- 98
 DB 808 TACAAATATTCATCTATTTTTAAAGAAAGCAATATTAACGATCCAGTGGTGGCTGCT 867
 QY 98 ----- 98
 DB 868 GCTCAAGATGATTCAGATGCTCTTCAAAAGAGATGCAATTACAGATGATGATG 927
 QY 99 -----ValAspGluValArgProLeuValTyr 107
 DB 928 GAAGTTCTGGCCGCGCATGATCTATATCAAGTTGAGGAGGATGAGGCTGATATAT 987
 QY 108 PheGluGlyThrLysAspSerProLeuGlnHisIleLysTyrValValSerTyrValAsn 127
 DB 988 TTGTAAGGACCAAGAGCTCCCTTTAGAGCATCTCTGATGATGATGATGATGATGAT 1047
 QY 128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
 DB 1048 CCTGGAGAGTGACAAAGGCTGACTACCGTACTACTACATTTCTGCTGATCAGTCAAG 1107
 QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAspProHisCysValSerLeu 167
 DB 1108 CACTGTGACTTTTATTAAGTATGATGATTAACCAAGAAATCCACACTGTGTGCTCTT 1167
 QY 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluIlePheTrpAlaThr 187
 DB 1168 TCAAGACTATCAAGTCCGTAAGATGACCAACTTGCAAAACAAAGAAATTTTGGGCAC 1227
 QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
 DB 1228 ATTTTGATTCAGCAGGCTCTCTCTGCTACTACTCTCCCAAGAAATTTCTTTTGAA 1287
 QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuIleProGly 227
 DB 1288 AGTACTACTGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1347
 QY 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGluValGlnLeuValAsn 247
 DB 1348 AAGAAATATCCACAGTGTGTTATATATGATGATGATGATGATGATGATGATGATGAT 1407
 QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 DB 1408 CGGTTTAAAGGAGTCACAGTATTTCCGTTGAATACCTTACGCTCTAGTTATGATGCTT 1467

```

OY 268 ValVal11leAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
DB 1468 GTAGTGAATGACCAACAGGGGATCCCTGTCAACGAGGCTTAATTGGAAGGCCCTTTAA 1527
OY 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuLys 307
DB 1528 TATTAATAAG----- 1536
OY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 327
DB 1536 ----- 1536
OY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
DB 1537 ----- GTTGTATTGCTGGGCC 1554
OY 348 ProValThrLeuThrPhePheTyrAspThrGlyTyrThr 360
DB 1555 CCACTGACTCTGTGATCTTCTATGATACAGGATACACG 1593
RESULT 12
AAD23843
ID AAD23843 standard; cDNA; 2510 BP.
XX
AC AAD23843;
XX
DT 07-MAR-2002 (first entry)
XX
DE Human protease PRTS-2 cDNA.
XX
KM Human; protease; PRTS-2; tranquiliser; gene therapy; vaccine; allergy;
KM infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;
KM atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;
KM gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;
KM epithelial disorder; urticaria; anorexia; trauma; asthma; eczema; nausea;
KM hypertension; neurological disorder; Parkinson's disease; drug screening;
KM cardiast; cell proliferative disorder; multiple sclerosis; osteoporosis;
KM diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;
KM autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;
KM developmental disorder; reproductive disorder; infertility; diarrhoea;
KM dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 616..2358
FT FT /*tag= a
FT FT /product= "Human protease PRTS-2 protein"
XX
PN MO200183775-A2.
XX
PD 08-NOV-2001.
XX
PF 04-MAY-2001; 2001MO-US14651.
XX
PR 04-MAY-2000; 2000US-202082P.
PR 11-MAY-2000; 2000US-203566P.
PR 17-MAY-2000; 2000US-205803P.
PR 25-MAY-2000; 2000US-207477P.
PR 01-JUN-2000; 2000US-209402P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Deleage AM, Lal P, Hafalia A, Patterson C, Walla NK, Kearney L,
PI Tribouley CM, Khan FA, Yao MG, Baughn MR, Azimzai Y, Elliott VS,
PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DM,
PI Reddy R, Yue H, Tang YT;
XX
DR WPI: 2002-034518/04.
DR P-PSDB; AAE14337.
XX
PT Novel human proteases and polynucleotides encoding the proteases,

```

PT useful for treating, diagnosing or preventing cell proliferative,
PT cardiovascular, autoimmune/inflammatory, neurological and developmental
PT disorders -

PS Claim 5; Page 139-140; 151pp; English.

CC The invention relates to human proteases (PRTS-14) and its corresponding
CC cDNA molecules. Human PRTS and its nucleic acid molecule are useful for
CC the diagnosis, treatment and prevention of disorders associated with
CC increased or decreased expression of PRTS. Examples of such disorders
CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,
CC hepatitis, psoriasis and cancer); autoimmune/inflammatory disorders
CC (AIDS, Addison's disease, allergy, anaemia, asthma, atopic dermatitis,
CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,
CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and
CC viral, bacterial, fungal, parasitic, protozoal and helminthic
CC infections); cardiovascular disorders (myocardial infarction, ischaemic
CC heart disease and hypertension); neurological disorders (epilepsy,
CC Alzheimer's disease, Pick's disease, Huntington's disease, dementia,
CC Parkinson's disease, stroke, mental disorders including mood, anxiety
CC and seasonal affective disorder and prion diseases); gastrointestinal
CC disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);
CC epithelial disorders (contact dermatitis, eczema, acne vulgaris,
CC alopecia, scabies, insect bites and urticaria); reproductive disorder
CC (infertility, disruption of estrous and menstrual cycle and
CC gynaecomastia); and developmental disorders (renal tubular acidosis,
CC Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).
CC PRTS DNA is also in useful in gene therapy. PRTS and its immunogenic
CC fragments are useful for screening libraries of compounds in several drug
CC screening assays. The present sequence is human protease PRTS-2 cDNA.

XX SQ Sequence 2510 BP; 777 A; 494 C; 527 G; 712 T; 0 other;

Alignment Scores:

Pred. No.: 8 78e-168 Length: 2510
Score: 1522.00 Matches: 309
Percent Similarity: 58.08% Conservative: 0
Best Local Similarity: 58.08% Mismatches: 0
Query Match: 78.74% Indels: 223
DB: Gaps: 2

US-10-070-464-7 (1-360) x AAD23843 (1-2510)

```

OY 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluLysPheAspArg 20
DB 609 GAAGAAATGCCAGATCGATGAGTGGCTACCTTGTCTCCAGAAAGAAATTGATGGA 668
OY 21 TyrSerGlyTyrTPTPTPCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
DB 669 TATTCTGCTATTGTGTGTGTCCAAAGCTCAAAACACTCCAGTGTGTAAATTTCTT 728
OY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleHisValIleThrSerPro 60
DB 729 AGAATTCATATGAGAAAATGATGAATCTAGGTGAAATTTATTCATGATACATCCCT 788
OY 61 MetLeuGluThrArg-ArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPr 80
DB 789 ATGTTGGAACCAAGCAGGCGAGATTCATTCCTGTTATCTAAACAGTCAACGCAATCC 848
OY 80 oLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
DB 849 TAAAGTCACTTTTAAAGATGTCAGAAATTAATGATGATGCTGAAGAAAGATCATGATGT 908
OY 98 ----- 98
DB 909 CATGATATGAAGAACTAATTCACCTTTTGAGATTCTATTGAAGAGCTGATATATTTC 968
OY 98 ----- 98
DB 969 CAGAGCTGATGACTCCTGAGGAAATATAGTTGTGTCATCTCTAATGATGCTGCCA 1028
OY 98 ----- 98

```

```

Db      1029 GACTCGCCTACAGATAGTGTGATCTCACCCTGAATTATTATCCAGTAGAAGATGATGT 1088
QY      98 ----- 98
Db      1089 TATGGAAGGACAGACTCATTTGATGTCAGTGCCTGATCTGTGAGCCACTAATTATCTTA 1148
QY      98 ----- 98
Db      1149 TGAGAAACACACAGACATCTGATTAATATCCATGACATCTTTGATGTTTCCCAAG 1208
QY      98 ----- 98
Db      1209 TCACGAGAGAAATTGAGTTTATTTTGGCTCTGATGCAAAACAGGTTCCGTCATTT 1268
QY      98 ----- 98
Db      1269 ATACAAAATTACCTATTATTTTAAAGAAACAAATATTAACGATCCAGTGGGCTGCC 1328
QY      98 ----- 98
Db      1329 TGTCCAACTGTCACTTGGATGATCAATTCATGAGATCTCTAGAACTCCATCCTGTAT 1388
QY      99 -----ValAspGluValArgArgLeuValTyrPheGlu 109
Db      1389 GTGTGTGACACATATATGTTGAGATCCAGTTGATGAGTCAGAAAGCGCTGATATTTTGA 1448
QY      109 uGlyThrIlyAspSerProleuGluIuhIshIleuTyrValIvalSerTyrValAsnProGlu 129
Db      1449 AGGACACCAAGACTCCCTCTTTAGACATCACTGTCGATGTCATGATGATGATGATGATG 1508
QY      129 yGluValIThrArgLeuThrAspArgIlyTyrSerHisSerCysCysIleSerGlnHisCys 149
Db      1509 AGAGGTGACAGAGGCTGACTGACCGTGGCTACTCACAATCTTGTGTGATCAGTCAGACAG 1568
QY      149 sAspPhePheIleSerIlyTyrSerAsnGlnIlyAsnProHisCysValSerLeuTyrIly 169
Db      1569 TGACTCTTTATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1628
QY      169 sleuSerSerProGluAspAspProThrCysIlyThrIlyGluPheTyrIle 189
Db      1629 GCTATTAAGTCTCGAAGATGACCCCACTTGCAAAACAAAGAAATTTGGCCACCACTTTT 1688
QY      189 uAspSerAlaGlyProleuProAspTyrThrProProGluIlePheSerPheGluSerTh 209
Db      1689 GAAATCAGCAGGCTCTTCTGACTATCTCTCTCAGAAATTTCTCTTTGAAAGTTC 1748
QY      209 rThrGlyPheThrLeuTyrGlyMetLeuTyrIlyProHisAspLeuGlnProGlyIlyIly 229
Db      1749 TACTGATTTTACCTTGTATGAGATGCTCTCAAGCCTCATGATCTACACCTGGAAGAA 1808
QY      229 sTyrProThrValIleuPheIleTyrGlyIlyProGluIvalGlnLeuValAsnAsnArgPh 249
Db      1809 AATCTCTACTGCTGCTTCAATATGTCGTCTCTCAGGTGACCTTGTAATATGCTT 1868
QY      249 eIlyGlyValIlyTyrPheArgLeuAsnThrIleuAlaSerLeuGlyTyrValValValVa 269
Db      1869 TAAAGAGTCAAGTATTTCCGCTTGATATACCTCAGCCTCTGAGTTATGTTGTTAGT 1928
QY      269 lIleAspAsnArgGlySerCysHisArgGlyLeuIlyPheGluGlyValAlaPheIlyTyrIly 289
Db      1929 GATAGCAACAGGGAGTCTGTACCGAAGGCTTAATTTGAAGGGCCCTTAAATATA 1988
QY      289 sMetGlyGlnIleGluIleAspAspIleValIleGluIlyLeuGlnTyrLeuAlaSerArgTyr 309
Db      1989 AATG----- 1992
QY      309 rAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyIlyTyrIleuSe 329
Db      1992 ----- 1992
QY      329 rIleuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProVa 349
Db      1993 -----GTTGCTATTGCTGGGGCCCCAGT 2015

```

```

QY      349 lThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
Db      2016 CACTCTGTGATCTTCTATGATACAGATACAG 2049
RESULT 13
ABK83332
ID      ABK83332 standard; cDNA; 4685 BP.
AC      ABK83332;
DE      12-AUG-2002 (first entry)
XX      cDNA encoding human DPRP-1 splice variant #8.
KW      Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW      DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW      diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW      heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW      ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW      dyskinesia; reproductive disorder; inflammatory disorder;
KW      metabolic disorder; gene; ss.
XX      Homo sapiens.
XX      OS
XX      PN      MO200231134-A2.
XX      PD      18-APR-2002.
XX      PF      12-OCT-2001; 2001MO-US31874.
XX      PR      12-OCT-2000; 2000US-240117P.
XX      PA      (FERR ) FERRING BV.
XX      PI      Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX      DR      WPI: 2002-444178/47.
XX      DR      P-P5DB; eABG61601.
XX      PT      New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
XX      PT      the proteins, useful for treating e.g. fungal, bacterial, protozoan and
XX      PT      viral infections, cancers, allergies, neurological disorders, or pain
XX      PS      Disclosure: Page 75-76; 113pp; English.
XX      CC      The present invention relates to the isolation of novel human serine
XX      CC      proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
XX      CC      proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
XX      CC      and nucleic acids encoding them are useful for treating infections
XX      CC      such as fungal, bacterial, protozoan and viral infections, particularly
XX      CC      infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
XX      CC      pain, diabetes, precocious puberty, infertility, obesity, anorexia,
XX      CC      bulimia, Parkinson's disease, acute heart failure, hypotension,
XX      CC      hypertension, urinary retention, osteoporosis, angina pectoris,
XX      CC      stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
XX      CC      psychotic and neurological disorders (e.g. anxiety, dementia, or
XX      CC      schizophrenia), and dyskinesias. These may also be used in discovering
XX      CC      therapeutic agents for the treatment of reproductive, inflammatory and
XX      CC      metabolic disorders. ABK83332-ABK83343 encode human DPRP proteins.
XX      SQ      Sequence 4685 BP; 1430 A; 853 C; 991 G; 1411 T; 0 other;
Alignment Scores:
Pred. No.: 2,26e-167 Length: 4685
Score: 1522.00 Matches: 312
Percent Similarity: 58.54% Conservative: 0
Best Local Similarity: 58.54% Mismatches: 1
Query Match: 78.74% Indels: 221
DB: 24 Gaps: 2
US-10-070-464-7 (1-360) x ABK83332 (1-4685)

```

QY 1 GIUGIUSPALAArgSerAlaGlyValAlaThrPheValLeuGlnGluIlePheSerPheGlu 20
 Db 943 GAAGAAGATGCCAGATCAAGCTGAGTGGCTACCTTTGTTCTCCAGAAAGAAATTTGATAGA 1002
 QY 21 TysSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
 Db 1003 TATTTGGCTATTTGGGTGTGTCCAAAAGCTGAAACAACTCCCAAGGTGTGTAATTCCT 1062
 QY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
 Db 1063 AGAATTCATATGAAAGAAATATGATGATCTGAGGTGGAATTTTCATCTTACATCCCT 1122
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
 Db 1123 ATGTTGAAACCAAGAGAGGCGCAGATTCATCCGTTATCTTAAACAGGTACCAATACCT 1182
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleLeu----- 98
 Db 1183 AAGTCACTTTTAAGATGTCAAGAAATATGATTTGATGCTGAAGAGATCATAGATGTC 1242
 QY 98 ----- 98
 Db 1243 ATAGATAAGAACTAATTCACTTTTGAGATTCTATTTGAAGAGATTGAATATATGTC 1302
 QY 98 ----- 98
 Db 1303 AGAGCTGATGACTCTGAGGAGAAATATGCTTGTCATCTACTAGATGCTCCAG 1362
 QY 98 ----- 98
 Db 1363 ACTGCCCTACAGATAGTGTGATCTCAGCTGAATATTATTCCTCAGTAGAAGATAGTT 1422
 QY 98 ----- 98
 Db 1423 ATGAAAGCAGAGACTCATTTGATGATGAGTCTGATTTCTGTGAGCGCACTATATCTAT 1482
 QY 98 ----- 98
 Db 1483 GAAAGAACACAGACATCTGATTAATATCCATGACATCTTTCATGTTTCCCAAGT 1542
 QY 98 ----- 98
 Db 1543 CACGAGAAGAAATTGAGTTTATTTTGCTCTGATGCAAAACAGGTTCCGTATTTA 1602
 QY 98 ----- 98
 Db 1603 TACAAATTTACATCTATTTTAAAGAAAGCAATATATAACATCCAGTGTGCTGCT 1662
 QY 98 ----- 98
 Db 1663 GCTCCAAAGTATTCAGTGTCTCTATCAAGAGAGATAGCAATTACAGTGTGATG 1722
 QY 99 -----ValAspGluValArgArgLeuValTyr 107
 Db 1723 GAAGTCTTGGCCGCGCATGATCTAATATCCAAAGTTGATGATGAGAGGTGGATATAT 1782
 QY 108 PheGluGlyThrLysAspSerProLeuGluHisIleLysValValSerTyrValAsn 127
 Db 1783 TTTGAAGGCAACCAAGACTCCCTTTAGAGCATCACTGTAGTGTAGTACCTAAT 1842
 QY 128 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyLeuSerGln 147
 Db 1843 CCTGAGAGGTGACAAAGCTGACTGACCGTGTCTACACATCTGTGCTGATCAGTCA 1902
 QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
 Db 1903 CACGTGTCTCTTTATAGTATAGTATAGTAAACGAAGATCCACATCTGTGTCCCTT 1962
 QY 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
 Db 1963 TACAAGCTATCAAGTCTGAAGATGACCAACTTCAAAACAAAGAAATTTTGGGCCAC 2022

QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
 Db 2023 ATTTGGATTTCAGT----- 2036
 QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
 Db 2036 ----- 2036
 QY 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGluValGlnLeuValAsnAsn 247
 Db 2037 -----CCTCAGGTGCAAGTTGGATATAT 2060
 QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 Db 2061 CGGTTAAAGGAGTCAGATATTCGCTGAATACCTTAGCTCTAGATTATGTGTT 2120
 QY 268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGluValPheLys 287
 Db 2121 GTAGTGAATGACAAACAGGAGATCCGTGACCAAGGCGCTTAAATTTGAAGCGCCCTTTAAA 2180
 QY 288 TyrLysMetGlyGluIleGluLeuAspAspGlnValGluGlyLeuGlnThrLeuAlaSer 307
 Db 2181 TATTAATGCGTCAATATGAAATTCACATCAGGTGGAAGAGCTCCAAATATCTAGTTCT 2240
 QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 327
 Db 2241 CGATATGATTTTCATGACTAGATGCTGTGGGCATCCACGCTGGCTTATGAGAGATAC 2300
 QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
 Db 2301 CTCTCCCTGATGCAATTAATGACAGGTCAGATATCTTCAGGAGGTGCTATTTGCTGGGCC 2360
 QY 348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
 Db 2361 CCAGTCACTCTGTGATCTTCTATGATACAGATACAG 2399
 Db
 RESULT 14
 ABRK83331
 ID ABRK83331 standard; cDNA; 4676 BP.
 AC
 AC ABRK83331;
 AC
 DT 12-AUG-2002 (first entry)
 XX
 XX cDNA encoding human DPRP-1 splice variant #7.
 DE
 XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinisia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder; gene; ss.
 XX
 XX Homo sapiens.
 OS
 PN WO200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 XX
 DR WPI, 2002-444178/47.
 DR P-PSDB; ABG61600.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 the proteins, useful for treating e.g. fungal, bacterial, protozoan and

PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -
 XX Disclosure, Page 72-73; 113pp; English.
 PS
 XX The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g., anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABK83322-ABK83343 encode human DPPR proteins.
 XX
 SQ Sequence 4676 BP; 1424 A; 859 C; 979 G; 1414 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,58e-167 Length: 4676
 Score: 1521.50 Matches: 309
 Percent Similarity: 57.97% Conservative: 0
 Best Local Similarity: 57.97% Mismatches: 1
 Query Match: 78.71% Indels: 224
 DB: 24 Gaps: 2
 US-10-070-464-7 (1-360) x ABK83331 (1-4676)
 QY 1 GIUGLHASPALAArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
 Db 943 GAAGAAGATGCCAGATCAAGCTGAGTCGCTACCTTGTCTCTCCAAAGAAATTTGATGA 1002
 QY 21 TYSerGlyTyrTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleu 40
 Db 1003 TATTCGGCTATGGTGGTGTCCAAAAGCTGAAACAACTCCAGTGGTGTAAATTTCTT 1062
 QY 41 ArgGlyLeuTyrGlnGluLysAspGluSerGluValGluIleIleHisValThrSerPro 60
 Db 1063 AGAATTCATATGAAGAAATGATGATCTGAGGTGAAATATCATGATCATCTCCCT 1122
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAspPro 80
 Db 1123 ATGTTGGAACAAGAGGGGCAATTCATTCCTTATCTTAAACAGGTCACAGAAATCTT 1182
 QY 81 LysValThrPheLysMetSerGlyIleMetIleAspAlaGlnGlyValGlyIleIle 98
 Db 1183 AAAGTCATTTTAAAGATGTCAGAAATATGATGATGATGATGATGATGATGATGATG 1242
 QY 98 ----- 98
 Db 1243 ATAGATAGAGAACTAATTCACCTTTTGAGATTCTATTGGAAGAGTGAATATATTTGCC 1302
 QY 98 ----- 98
 Db 1303 AGAGCTGATGACCTCTGAGGGAATAATGCTTGGTCATCTCTACTAGATGCTCCAG 1362
 QY 98 ----- 98
 Db 1363 ACTCGCTTACAGATAGTGTGATCTGCACCTGAATTTATTTATCCAGTAGAAGATGAT 1422
 QY 98 ----- 98
 Db 1423 ATGGAAGGACAGAGACTCATGTAGTACGCTGATTTCTGTAGCCCACTAATATCTAT 1482
 QY 98 ----- 98
 Db 1483 GAAGAAACAACAGACATCTGATTAATATCATGATCATCTTTTCCCAAGT 1542
 QY 98 ----- 98

Db 1543 CACGAAGAGAAATGAGTTTATTTTGGCTCTGATGCAAAACAGTTTCGTCATTTA 1602
 QY 98 ----- 98
 Db 1603 TACAAAATATCATCTATTTTAAAGAAAGCAATATAAACGATCCAGTGGGCTGCTCT 1662
 QY 98 ----- 98
 Db 1663 GCTCCAGATGATTTCAAGTGTCTTCAAAAGAGAGATAGCAATTTCCAGTGTGAATGG 1722
 QY 99 -----ValAspGluValArgArgLeuValTyr 107
 Db 1723 GAAGTCTTGGCGGCATGATCTATATCCAAAGTTGAAGAGCTCAGAAAGCTGATAT 1782
 QY 108 PheGluGlyThrLysAspSerProLeuGluHisIleLysTyrValValSerTyrValAsn 127
 Db 1783 TTGGAAGGACCAAGAGACTCCCTTTAGAGCATCACTGTACGTAGTACGTAAAT 1842
 QY 128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
 Db 1843 CCTGAGAGGTGACAAAGGCTGACTGACCTGCTACTCATCTTCTGTCATCACTGACG 1902
 QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
 Db 1903 CACTGAGACTCTTTATATAGTATAGTATAGTAAACAGAAATCCACAGCTGTGCTCCTT 1962
 QY 168 TyrLysGluSerSerProGluAspAspProThrCysLysThrLysGluThrPalaThr 187
 Db 1963 TACAAGCTATCAAGCTCGAAGATGACCCAACTTCCAAAACAAAGAAATTTTGGGCAAC 2022
 QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGln 207
 Db 2023 ATTTGGATTCAGCAGGCTCTCTCTGCTATACCTTCCACGAAATTTCTTTTGA 2082
 QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
 Db 2083 AGTACTACTGATTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2142
 QY 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
 Db 2143 AAGAAATATCCATCACTGCTGCTCATATATGATGATGATGATGATGATGATGATG 2180
 QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuLysSerLeuGlyTyrValVal 267
 Db 2180 ----- 2180
 QY 268 ValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
 Db 2180 ----- 2180
 QY 288 TyrLysMetGlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnIleLeuAlaSer 307
 Db 2181 -----GTCAAATAGAAATGACGATCAGGTGGAAGACTCCAAATATCTAGGTTCT 2231
 QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 327
 Db 2232 CGATATGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2291
 QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
 Db 2292 CTCTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2351
 QY 348 ProValThrLeuThrIlePheTyrAspThrGlyTyrThr 360
 Db 2352 CAGTCACTCTGTGATCTTCTATGATGATGATGATGATGATGATGATGATGATGATG 2390
 RESULT 15
 ID ABV76411
 XX ABV76411 standard; cDNA; 1821 BP.
 AC ABV76411;
 XX

Db	2486	GGCGGCGCGGACTGGATGCTATGAAGAAGAGATCCCGCGGATCGCTAACGC	---TTGNG	2542
Qy	26	TrycysProlysalagluThrThrProserglyglylysileuarglleuTyrglu	45	
Db	2543	TTTCGCGCAGATGGTGAA	-----TATCGGCTTACTG	2575
Qy	46	GluaenaePgluSerGluValGluIlelleHisValThrserProMetIeuGluThrarg	65	
Db	2576	AGCTTCATGAGACCTGGGGTTCCGACCTACACCGTTCAAGTATATATGATTAACCAAG	2635	
Qy	66	ArgAlaasPserPhe	-----ArgTyrrProlyserThrGlyThrAlasn	79
Db	2636	ATCGCTCCGGCGGTATCCATGAGAGCTGAGACTGAGATTAAGGTATCCCAAGGTTCGCACAGCAAT	2695	
Qy	80	ProlyValThr	-----	83
Db	2696	CCGACCGTGAAGTTGATGCTGCTTACATCGCTACAGAGAGTGAGCAAGCGCCGATC	2755	
Qy	84	-----PheylsMetSerGluIleMetIle	-----	91
Db	2756	GACGGGTTCGAGTCACTGACTGATCATTTGGCGAGGTGCTGGCTCACTGATCTAC	2815	
Qy	91	-----	-----	91
Db	2816	ACCACCGTGTGCTTAAGCGCTTCAACCGTGTCCAGACCAAGAAAGTCGTGGGTC	2875	
Qy	92	-----AspAlaGluGlyArg	-----	96
Db	2876	GATACTGCGCTGGAAACAAGGCTACTGTCTATCAGACCAAGATGGACCGATGATGCTC	2935	
Qy	96	-----	-----	96
Db	2936	GATACCTCTTTCATGATGATATATGACCTTCAAGCGGTCCGACAAAGATGCTAC	2995	
Qy	96	-----	-----	96
Db	2996	TACATCGACATCTCGACCATTCGGGATGGGGGATCTGTATCTTCCCGGTTTCGGGC	3055	
Qy	97	-----IleIleValAsp	100	
Db	3056	GCGCAACTATCCCACTAACCAAGAAGCGACTGGAGATCAAGTATATCTGAGTATGAT	3115	
Qy	101	GluValArgArgGluValTyrrPheGluGlyThrIlysaPserProIeuGluHisIleu	120	
Db	3116	CAGGAACCCCAAGTGGTGTATCTACTCTGCTGACTCAACCAACAGACCGAGCGCCATCTC	3175	
Qy	121	TyrValValSerTyrrValAsnProGlyGluValThrArgIeuThrAspArgIlyTyrrSer	140	
Db	3176	TACTCGCTCTCTAT	---TCCACGTTGGGGTCAACCGGCTCTGTGAGACGACCGTGTGC	3232
Qy	141	-----HisSerCysCysIleSerGlnHisCysAspPhePheIleSerIlyTyrrSerAsn	158	
Db	3233	GCGTACTGCTGCTTCTCTCTCCGCGAAGCTGGGCTACTCATCTCACATACGAGAC	3292	
Qy	159	GlnIlysaSnProHisCysValSerIeuTyrrIlylsIeuSerSerProGluAspAspProThr	178	
Db	3293	CCAGACGTAACCTTAC	---CAGAACTTCAACGACCAACAGT	3331
Qy	179	CysIlyThrIlyGluPheThrAlaThrIleIeuAspSerAlaGlyProIeu	195	
Db	3332	-----ACCAACAACAATC	---CGCAATATCAACGCAACGCGCAAAAGTCTCGAGCAAAATC	3382
Qy	196	ProAspTyrrThrProProGluIlePheSerPheGlu	-----SerThrThrGlyPheThr	213
Db	3383	AAGACATATGATTCGCCAACATCACTACTTCAGACTTCCCTCCCTCCGAGAAACC	3442	
Qy	214	LeuTyrglyMetLeuTyrrIlyserProHisAspIeuGlnProGlyIlyIlyTyrrProThrVal	233	
Db	3443	CTCATATGATGACAGCGCTTACCCCGGGTTCCTCCCGAGTAAAGAACTACCCCATACTT	3502	
Qy	234	LeuPheIleTyrglyGlyIlyProGluValGlnIleuValaIasnAsnArgPheIlyGlyValIlys	253	
Db	3503	TTGACCCCATACGGCGGCCCAAGCGGCCCAAGATGATCAACAAGATGGCAAGCCCTGAT	3562	

QY	254	Tyr--PheArgLeuAenThleuAlaSerLeuGlyTyrValValValValIleAspAsn	272
Db	3563	TTCAAGGCTTAATGTCGCCCTCCGACGACGAACTCCAGTACCTGACCTGTGCACAAAC	36222
QY	273	ArgGlySerCysHisArgGlyLeuLeuPheGluIuLysAlaPheLysTyrLysMetGlyGln	292
Db	3623	CGCGGACACAGGTTTCAAAGAGCGCAAGTTCGCCCTCCGCCCTGCACGCCCACTGGGCTC	3682
QY	293	IleGluIleAspAspGlnValGluGlyLeuGlnIleTyrLeuAlaSerArgTyrAspPheIle	312
Db	3683	CTCGAAGACAGAAAGCACGATCTACGCGCGGCAACAG--GCGGCAACATCCCTCGATC	3739
QY	313	AspLeuAspArgValGlyIleHisGlyTyrPseTyrGlyGlyTyrLeuSerLeuMetAla	332
Db	3740	GATCCACAGCCACCTGGGATCTGGGGCTCGAGGTTTTCGGAGGCTACTTGACCACGACAGCTC	3798
QY	333	LeuMetGlnAagSerAspIlePheArgValAlaIleAlaGlyAlaProValThleuTyr	352
Db	3800	CTGGAGAAAGACACCGGTGCTTTACATTAGAGATCATCACCGGCCCTGGTTTCTGCATCGG	3855
QY	353	IlePheTyrAspThrGlyTyrThr	360
Db	3860	CGTTTCTACGACTCAATGTACACG	3883

```

RESULT 2
US-09-280-116-171
; Sequence 171; Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 171
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: prolyl oligopeptidases
US-09-280-116-171

```

Alignment Scores:	
Pred. No.:	1.09e-11
Score:	346.00
Percent Similarity:	98.57%
Best Local Similarity:	95.71%
Query Match:	17.90%
OB:	4
Length:	822
Matches:	67
Conservative:	2
Mismatches:	0
Indels:	0
Gaps:	1

US-10-070-464-7 (1-360) x US-09-280-116-171 (1-823)

QY	252	GlnIleGluIleAspAspGlnValGluGly-LeuGlnIleTrpLeuAlaSerArgTyrAsp	311
Db	3	CAAAATGAAATTTGACGATCAGGTGGAGAGCAATCCAAATATCTAGCTTCGATATATAT	62
QY	311	eIleAspLeuAspArgValGlyIleHISglYrPserTyrGlyGlyTyrLeuSerLeu	331
Db	63	CATTACTTGAATCGTGTGGCATCCACGGCTGGTCTTATGAGAGATACCTCTCCGTAT	122
QY	331	AlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThr	351
Db	123	GGCATTATATGACAGAGTCAGATATCTTCAGGGTGTCTATGTGCTGCCCACTACTCT	182
QY	351	uTrpIlePheTyrAspThrGlyTyrThr	360
Db	183	GTGATCTTCTATGATACAGGATAACG	210

RESULT 3
PCT-US93-07923-1

QY 308 ArgTYrAspPheIleAspLeuAspArgValGlyIleHISglYTrpSerTYrGlyGlyTYr 327
DB 1853 AAAATGGAGATTGTGGACACAAACAAATTGCAATTGGCGCTGGTCATATGGAGGGTAC 1912
QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
DB 1913 GTAACCTCAATGGCTCGGTGAGTACAGGAAGTGGCGTTCATCAAGTGGAAATAGCCCGGGG 1972
QY 348 ProValThrLeuThrIlePheTYrAspTYrGlyTYrThr 360
DB 1973 CCTGATCCCGGTGGAGTACTATGACTAGCTAGTACACA 2011

RESULT 4
US-10-002-593-5
Sequence 5, Application US/10002593
Patent No. 6586198
GENERAL INFORMATION:
APPLICANT: Vanderbilt University
APPLICANT: Brown, Nancy J.
TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTIN
FILE REFERENCE: Acty Docker No. 6586198 1242/48/2
CURRENT FILING DATE: 2001-10-31
PRIOR APPLICATION NUMBER: US/10/002,593
PRIOR FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 3407
TYPE: DNA
ORGANISM: Homo sapiens
US-10-002-593-5

Alignment Scores:
Pred. No.: 5.08e-27 Length: 3407
Score: 315.00 Matches: 120
Percent Similarity: 34.89% Conservative: 59
Best Local Similarity: 23.39% Mismatches: 146
Query Match: 16.30% Indels: 188
DB: 4 Gaps: 22

US-10-070-464-7 (1-360) x US-10-002-593-5 (1-3407)

QY 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValIleGlnGluGlu---Pheasp 19
DB 643 AAAGAAGATATATATATATATGGAATGACTGGGTTTATGAAAGAGAAAGCTTTCAGT 702
QY 20 ArgTYrSerGlyTYrTYrTYrPCysProGlySalGluThrThrProSerGlyGlyIle 39
DB 703 GCCTACTCTGCTCTGTGGGTGCTCCAAACGGCACTTT----- 741
QY 40 LeuValGlyLeuTYrGlyGluGlnAsnAspGlnSerGluValGluIleIleHis----- 56
DB 742 -----TTAGCATATGTCCCAATTTTAAAGACAGAGAGTCCCACTTATTTGAATATCTCCCTTC 795
QY 57 ValThrSerProMetLeuGlnThrArgAlaAspSerPheArgTYrProIleThrgly 76
DB 796 TACTCTGATGAGTACTGAGTACCCAAAGACTGTACGGGTTCCATATCCAAAGCAGAGA 855
QY 77 ThrAlaAsnProIleValThrPhe----- 84
DB 856 GCTGTGAATCCCAACTGTAAAGTTCTTTGTGTAATACAGACTCTCTCAGCTCAGTCACACC 915
QY 85 LysMetSerGluIleMetIleAspAlaGlnGlyArgGlyIleIleValAspGluValAlaArg 104
DB 916 AATGCAACTTCATACAAATCACTGCTCTCTCATGTGATAGG----- 963
QY 105 LeuValTYrPheGlnGlyThrLysAspSerProLeuGlnHisIleuTYrValAlaSer 124
DB 964 -----GATCACAACCTTGTGATGAGACA 987
QY 125 TyrValAsnProGlyGluValThr-----ArgLeuThrAspArgGlyTYr 139

DB 988 TGGCGAACACAGAAAGAAATTTCTTTCAGTGGCTCAGAGAGATTCAGAACTATTCGTC 1047
QY 140 SerHisSerCys-----CysIle-----SerGln 147
DB 1048 ATGATATATTGTGACTATGATGATATCCAGTGGAGATGGAATCGCTTAGTGACCGGCAA 1107
QY 148 HisCysAspPhe-----PheIleSerLysTYrSerAsnGlnLysAsnProHis 163
DB 1108 CACATTGAAATGACTACTGCTGGCTGGTGAAGATTT--AGGCTTCAGAACTCAT 1164
QY 164 CysVal-----SerLeuTYrLysLeuSerSerProGluAsp----- 175
DB 1165 TTACCCCTTGATGTAATAGCTTCTTCAAGATCATCAGAAATGAGAAAGATTACAGACAC 1224
QY 176 -----AspProThrCysLysThrLysGluPheTrpAla 186
DB 1225 ATTGCTATTTCCAAATAGATATAAAGACCTCAGATTTATTAACAAAGCACCCTGGAA 1284
QY 187 ThrIle----- 188
DB 1285 GTCATCGGATGAAAGCTCTTAACAGTATTATCTACTACTATTAAGTAATATAA 1344
QY 188 ----- 188
DB 1345 GGATGCCAGAGAGAAAGAAATCTTATAAAATCCAATTATGACTATACAAAGTGACA 1404
QY 188 ----- 188
DB 1405 TGCCCTCAGTTGTAGCTGATCCGAAAGGTCTAGTACTATTCTGTGCAATTCAGTAA 1464
QY 189 -----LeuAspSerAlaGlyPro---LeuProAspTYrThr 199
DB 1465 GAGCGGAAGTATTATCACTGATGAGATGTTCCGATCTGCTGCTGCTCTTAATCTACAC 1524
QY 199 ----- 199
DB 1525 AGCAGCGTAATGATTAAGCGCTGAGAGTCCGTAAGAACAAATTCAGCTTTGATTAATAG 1584
QY 200 -----ProProGluIlePheSerPhe-----GluSerThrThrgly 211
DB 1585 CTGCAGAAATGCCAGATGCCCTCCCAAAAACCTGACTTATTTGATGATAAACAA 1644
QY 212 PheThrLeuTYrGlyMetLeuTYrLysProHisAspLeuGlnProGlyLysTYrPro 231
DB 1645 TTT---TGTATCAATGATATCTTCCCTCTCAT---TTGATTAATCCAAAGAAATATCT 1698
QY 232 ThrValLeuPheIleTYrGlyGlyProGlnValGlnLeuValAsnAsnArgPheIleGly 251
DB 1699 CTACTATTAGATGTATGACAGGCCCATGTACTCA-----AAAGCA 1740
QY 252 ValLysTYrPheArgLeuAsn-----ThrLeuAlaSerLeuGlyTYrValValAla 268
DB 1741 GACACTGTCTTCAGCTGAATCTGGGCCACTTACCTTGCAAGACAGAAAACATTATAGTA 1800
QY 269 Val---IleAspAsnArgGlySerCysHisArgGlyLeuLysPheGlnGlyIleAlaPheLys 287
DB 1801 GCTACTTGTATGAGAGAGAAAGTGTATACCAAGAGATAGATCATGATCATCAAC 1860
QY 288 TyrLysMetGlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnTYrLeuAlaSer 307
DB 1861 AGAAGACTGGGAACATTTGAAGTTGAAAGTCAAAATTTGAAGCAGCCACAATTT---TCA 1917
QY 308 ArgTYrAspPheIleAspLeuAspArgValGlyIleHISglYTrpSerTYrGlyGlyTYr 327
DB 1918 AAAATGGAGATTGTGGACACAAACAAATTGCAATTGGCGCTGGTCATATGGAGGGTAC 1977
QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
DB 1978 GTAACCTCAATGGCTCGGTGAGTACAGGAAGTGGCGTTCATCAAGTGGAAATAGCCCGGGG 2037
QY 348 ProValThrLeuThrIlePheTYrAspTYrGlyTYrThr 360

Db 2038 COTGTATCCGGTGGAGTACTATGACTCATGCTACACA 2076

RESULT 5

US-09-221-017B-646/c

Sequence 646, Application US/09221017B

Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: ROSE, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Montoy, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 646:

SEQUENCE CHARACTERISTICS:

LENGTH: 657 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: UNKNOWN

ORIGINAL SOURCE:

ORGANISM: PORYPHYROMONAS GINGIVALIS

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1...657

US-09-221-017B-646

Alignment Scores:

Pred. No.: 2.45e-24 Length: 657

Score: 283.00 Matches: 60

Percent Similarity: 55.21% Conservative: 30

Best Local Similarity: 36.81% Mismatches: 69

Query Match: 14.64% Indels: 4

DB: 4 Gaps: 2

US-10-070-464-7 (1-360) x US-09-221-017B-646 (1-657)

QY 198 TyrThrProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMet 217

Db 566 TACAACCCGAAGAGATTTCACCTATCAAACTCAATGCGCTTGAACTGATGCTCG 507

QY 218 LeuTyrLysProHisAspLeuGlnProGlyLysTyrProThrValLeuPheIleTyr 237

Db 506 ATCGTGAAGCTATTGATTTCGATCCCTCTGCGCACTATCCGTCTGATGTACAGTAT 447

QY 238 GlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeu 257

Db 446 AGCGGTCCCACTCCACAGCTATTTGATGCTAT-----TCATTCGATTGGGAA 396

QY 258 AsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgLysTyrCysHis 277

Db 395 CACTACCTTGATCGAAAGGTTTACCTCGTCGATGTCGATGTCGATGTCGATGTCG 336

QY 278 ArgGlyLeuLysPheGluGlyValAlaPheLysTyrLysMetGlyGlnIleGluIleAsp 297

Db 335 CCGGCGAAGATGCGCGAAGTGTACCTACATGCACTCGTGTATTGAAAGCGATGAT 276

QY 298 GlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgVal 317

Db 275 CAGATAGCAGCGGCCACTGCTAT---GACAGCTGCCCTATGTGATGACAGCTGATC 219

QY 318 GlyIleHisGlyTyrPheTyrGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSer 337

Db 218 GGCATATGGGGGTGATGCGCGCTATACCACTAATGATTTGTGTGGGGGAAT 159

QY 338 AspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThr 357

Db 158 GGTACATTCAAGCGGGATATAGCCTTGTCTCTGTGGAGACTGCGCTTTCACGATTCG 99

QY 358 GlyTyrThr 360

Db 98 GTTTCACCC 90

RESULT 6

US-08-230-491A-1

Sequence 1, Application US/08230491A

Patent No. 5587299

GENERAL INFORMATION:

APPLICANT: Retzig, Wolfgang J.; Scanlan, Matthew J.;

APPLICANT: Garth-Chessa, Pilar; Old, Lloyd J.

TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR

TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES

TITLE OF INVENTION: THEREOF

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESSES:

ADDRESSEE: FELFE & LYNCH

STREET: 805 THIRD AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: WORDPERFECT - ASC II

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/230,491A

FILING DATE: 20-APRIL-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 5587299man D.

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: IUD 330

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 688-9884

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2835 base pairs

TYPE: nucleic acid

```

: STRANDEDNESS: double
: TOPOLOGY: linear
US-08-230-491A-1

Alignment Scores:
Pred. NO.: 3,05e-20 Length: 2815
Score: 257.00 Matches: 114
Percent Similarity: 36.70% Conservative: 64
Best Local Similarity: 23.51% Mismatches: 158
Query Match: 13.30% Indels: 150
DB: Gaps: 20

US-10-070-464-7 (1-360) x US-08-230-491A-1 (1-2815)

QY      8 GlyValAlaIthrPheValLeuGlnGluGluPhe-----AspArgTyrSerGlyTyrTyr 25
DB      791 GGAATCCCAAGACTGTTTATGCAAGAGAAATGCTTCCTACAAATATAGCT---CTCTGG 847
QY      26 TrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrGlu 45
DB      848 TCGTCTCCT-----AATGAAAAATTTTTC-----GCATATGCG 880
QY      46 GluAsnAspGluSerGluValGluIleIleHisValThrSerProMetLeuGluThr--- 64
DB      881 GAATTTATGATGATAGAGATATACCAAGTATTCCTATTCCTATTATGCGGATGAACAATAT 940
QY      65 ArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnProLysValThrPhe 84
DB      941 CCTAAACAATAAATATATTCATATCCCAAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTA 1000
QY      85 LysMet-----SerGluIleMetIleAspAla 93
DB      1001 TTTATATATGATACCACTTACCCCTGATGATAGTCCCAAGAGTGGCTGTTCACAGTA 1060
QY      94 -----GluGlyArgIle 97
DB      1061 ATGATAGCCTCAAGTATATATTTCAAGTTCAGTTCAGCTGAGGTTACTGATGAACAGATA 1120
QY      98 IleValAspGluValArgArgLeuValTyrPheGlu----- 109
DB      1121 TGTTCAGTGGCTAAAGAGAGTCCAGAAATGTTGGCTCTCTCATATATGACCTTCAG 1180
QY      110 -----GlyThrLysAspSerPro--LeuGluHisLysLeuTyr----- 121
DB      1181 GAAGACTGGCAGACATGAGATTGTCCAAAGACCCAGAGCATATGAGAAGAAAGACAGACT 1240
QY      122 -----ValValSerTyrValAsnProGlyGlu 130
DB      1241 GGATGGGCTGTGATTCCTTTGTTTCAAGACCAAGTTTTCAGCTATGATGCCATTTCTGATC 1300
QY      131 ValThrArgLeuThrAspArg--GlyTyrSerHisSerCysCysIleSerGlnHisCys 149
DB      1301 TACAAAATATTTTATGAGCAAGAGATGGCTTCAAAACATAT--TCACATATCAAAAGACACTGT 1359
QY      150 AspPhePheIleSerLysTyrSerAsnGln----- 159
DB      1360 GGAATATGCTATCAAAATATCAAGTGGCAAGTGGAGGCCATATAATATATTCAGATAAC 1419
QY      160 -----LysAsnProHis 163
DB      1420 ACAGATTTCATGTTTATTTCTAGCAAGAAATTTGAAGAAATACCTCGAAGAAAGAAACT 1479
QY      164 CysValSer--LeuTyrLysLeuSerSerProGluAsp--AspProThrCysLysThrIly 182
DB      1480 CTACAGAAATTTGATGAGAGCTATCTCCCAAGCAAGAAAGTGTGTTACTTGCACATCTAAG 1539
QY      182 sGlu-----PheTyrAlaThrIleLeuAspSerAla----- 192
DB      1540 GAAAGAAAGTGCACATATATTCACAGCAAGTTCAGCGACTACGCGCAAGTACTATGCACT 1599
QY      193 -----GlyPro--LeuProAspTyrThr----- 199
DB      1600 TGTCTGCTACGGCCAGGCAATCCCATTTTCACCCCTTCATATGATGAGCGCATGATCAAGA 1659

```

```

QY      200 -----Pr 200
DB      1660 AATTAATCTCTGAAGAAAACAAGAAATGGAATGCTTTGAAAAATATCCAGCTGCC 1719
QY      200 oProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLy 220
DB      1720 TAAAGAGAAATTAAGAAACTTGAAGATGATGAATTAATTAATCTTATGATGCAAGATGATCT 1779
QY      220 sProHisAspLeuGlnProGlyLysTyrProThrValLeuPheIleTyrGlyGlyPr 240
DB      1780 TCTCTCTCAATTTGACAGATCAAGAAAGTATCCCTTGCTAATTCAGAGTATGATGGTCTCC 1839
QY      240 cGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLe 260
DB      1840 CTGACAGTCAAGAGTGAAGTCT-----GATTTCGTGTTAATTTGAT 1881
QY      260 uAlaSerLeu-----GlyTyrValValValIleAspAsnArgLysSerCy 276
DB      1882 ATCTTATCTTCCAGATGAAGAAAGGATGCTATGCTGCTGATGATGCTGATGATGATGATGAT 1941
QY      276 shiArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluHis 296
DB      1942 TTTCCAAAGTGAACAACTCTCTATGCAAGTATGCAAGAGCTGGGCTGTTTATGAAGTTGA 2001
QY      296 pAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspAr 316
DB      2002 AGACCAAGTTCACAGCTGTCAGAAATTCATA---GAATGGCTTCATTTGATGAAGAAAG 2058
QY      316 gValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnAr 336
DB      2059 AATGCAATATGAGGCTGTCTCTATGAGATAGCTTTCATCAGCTGAGCTGCTGATCTGCTG 2118
QY      336 sSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAs 356
DB      2119 AACTGCTCTTTTCAATGTGATATAGCAGTGGCTCCAGCTCCAGCTCCAGCTGGAATATTACGC 2178
QY      356 pThrGlyTyrThr 360
DB      2179 GTCTGTCTACACA 2191

RESULT 7
US-08-619-280A-1
; Sequence 1, Application US/08619280A
; Patent No. 5767242
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Reftig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
; TITLE OF INVENTION: ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Felte & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619, 280A
; FILING DATE: 18-MARCH-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5767242man D.
; REGISTRATION NUMBER: 30,946

```


FILING DATE: 01-OCT-1997
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/619,280
 FILING DATE: 18-MARCH-1996
 APPLICATION NUMBER: 08/230,491
 FILING DATE: 20-APRIL-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Hanson, No. 5965373man D.
 REGISTRATION NUMBER: 30,946
 REFERENCE/DOCKET NUMBER: LUD 5330.1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 688-9200
 TELEFAX: (212) 838-3884
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2815 Base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-940-391-1

Alignment Scores:
 Pred. No.: 3,05e-20 Length: 2815
 Score: 257.00 Matches: 114
 Percent Similarity: 36.70% Conservative: 64
 Best Local Similarity: 23.51% Mismatches: 158
 Query Match: 13.30% Indels: 150
 Gaps: 20

US-10-070-464-7 (1-360) x US-08-940-391-1 (1-2815)

QY 8 G1ValAlaThrPheValLeuGlnGluPhe-----AspArgTyrSerGlyTyrTrp 25
 Db 791 GGAATCCAGCTGGGTTTATGAGAGAAATGCTTCTCAAAATATGCT--CTCTGG 847
 QY 26 TrpCysProLysIaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrGlu 45
 Db 848 TGGCTCTCT-----AATGAAATTTTGG-----GCATTATGCG 880
 QY 46 GluAsnAspGluSerGluValGluIleIleHisValThrSerProMetLeuGluThr--- 64
 Db 881 GAATTTATGATAGATATACAGATATACAGATTATGCTATTCATATATGCGCATGAAACATAT 940
 QY 65 ArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnProLysValThrPhe 84
 Db 941 CCTAGAACATTAATATTCATACCAAGGCTGAGCTAAGATCCCGTTGTCGAGATA 1000
 QY 85 LysMet-----SerGluIleMetIleAspAla 93
 Db 1001 TTTATTTATCATACCACTTACCTGCGTATGATAGTCCCAAGAGATGCTGTTCCAGCA 1060
 QY 94 -----GluGlyArgIle 97
 Db 1061 ATGATAGCTCAAGTATATTTATTTCACTGCTCAAGCTGATTAAGTAAAGAGATA 1120
 QY 98 IleValAspGluValaGArgLeuValTyrPheGlu----- 109
 Db 1121 TGTTCAGTGGCTTAAAGAGTCACAGATGTTTGGCTCTGTCTATATGTGACTTCAGG 1180
 QY 110 -----GlyThrLysAspSerPro---LeuGluHisIleTyr----- 121
 Db 1181 GAAGACTGGCAGATAGTGGATTTCCAAAGACCCAGAGCATATAGAAAGAAAGAGAACT 1240
 QY 122 -----ValValSerTyrValAsnProGlyGlu 130
 Db 1241 GGAATGGCTGGTGGATTTCTTTGTTTCAAGACCACTTTTCAAGTATGATTCGATTCGAC 1300
 QY 131 ValThrArgLeuThrAspArg---GlyTyrSerHisSerCysGlyIleSerGluHisCys 149
 Db 1301 TACAAATATTTATGATGACAAAGATGGCTACAAACATAT-TCACTATATCAAAAGACCTGT 1359
 QY 150 AspPheHeIleSerLysTyrSerAsnGln----- 159

Db 1360 GGAATAATGCTATTCAAATTAACAAGTGGCAAGTGGAGGCCCATTAATATATTCAGATAC 1419
 QY 160 -----LysAsnProHis 163
 Db 1420 ACAGATTCAGCTGTTTATTTATTCAGCAATGAATTTGAAGATACCTCGTAAGAAAGAAACAT 1479
 QY 164 CysValSer---LeuTyrLysLeuSerSerProGluAsp-AspProThrCysLysThrLys 182
 Db 1480 CTACAGAAATTAAGATTGAAGGCTATCTCCAAAGCAAGAGTGTATTCCTGATTCAG 1539
 QY 182 sglu-----PheTrpAlaThrIleLeuAspSerAla----- 192
 Db 1540 GAAAGAAAGGTGCCAATATTACACAGCAAGTTTCAGCGACTACGCCAAGTACTATGCACT 1599
 QY 193 -----GlyPro---LeuProAspTyrThr----- 199
 Db 1600 TGTCTGCTACGGCCAGGATCCCATTTCCACCTTCATGATGACGACGATCAAGA 1659
 QY 200 -----Pr 200
 Db 1660 AATTAAATCTCGAAGAAACAGAAATTGAAATGCTTGAATAATATTCAGCTGCC 1719
 QY 200 oProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 220
 Db 1720 TAAAGGAAATTAAGAACTTGAAGTATGAATTAATCTTATGATGACAAAGATGATCT 1779
 QY 220 sProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPr 240
 Db 1780 TCCTCTCATATTGACAGATCAAGAAAGTATCCCTGCTAATCAAGTATGATGTCCTC 1839
 QY 240 oGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLe 260
 Db 1840 CTCAGCTCAAGTGAAGTGGTCT-----GATTTCTGTATTAATGGAT 1881
 QY 260 uAlaSerLeu-----GlyTyrValValValIleAspAsnArgLysSerCys 276
 Db 1882 ATCTTATCTTGGCAAGTAAAGAGAGATGCTATTCCTTGTGATGCTGAGAGAAAGC 1941
 QY 276 sHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGlyLeas 296
 Db 1942 TTTCGAAGTGACAAATCTCTTATGACAGTATTCGAAGATGCTGCTTTATGAAAGTTGA 2001
 QY 296 pAspGlnValGluGlyLeuGlnIleLysLeuAlaSerArgTyrAspPheIleAspLeuAspAr 316
 Db 2002 AGACCAAGATTACAGCTGTCAAGAAATTCATA---GAAATGGCTTCAATTGATGAAAAAG 2058
 QY 316 gValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnAr 336
 Db 2059 AATAGCATATGGGGCTGCTGCTTATGAGCATACGTTTCACTGCGCTTGATCGATCGG 2118
 QY 336 gSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAs 356
 Db 2119 AACTGGCTTTTCAAAATGTGTATAGCAGTGGCTCCAGTCCAGCTGGGAATTTATGCGC 2178
 QY 356 pThrGlyTyrThr 360
 Db 2179 GTCTGTCTACACA 2191

RESULT 9
 US-09-221-017B-253
 Sequence 253, Application us/09221017B
 Patent No. 6444799
 GENERAL INFORMATION:
 APPLICANT: Ross, Bruce C.
 TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
 NUMBER OF SEQUENCES: 1120
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORRISON & FOERSTER
 STREET: 755 PAGE MILL ROAD
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA

```

; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221.017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P11182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P11546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P12811
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 253:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...543
; US-09-221-017B-253

Alignment Scores:
Pred. No.: 6,286-19 Length: 543
Score: 236.50 Matches: 41
Percent Similarity: 65.00% Conservative: 24
Best Local Similarity: 41.00% Mismatches: 34
Query Match: 12.23% Indels: 1
DB: Gaps: 1

US-10-070-464-7 (1-360) x US-09-221-017B-253 (1-543)
QY 260 LeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCysHisArgGly 279
DB 43 ATGCACAGAAAGGCTATGCGCTTACGGGTATGATCGCGGATCGCCATGAGAGGG 102
QY 280 LeuIysPheGluGlyAlaPheLysTyrIysMetGlyGlnIleGluIleAspAsnVal 299
DB 103 GCTGCTTCGAGCAGGATTATTCATGCTGTTGGGCGAGACGAGATGCGCATGATG 162
QY 300 GluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 319
DB 163 TCGCGTGTGATTTCTCAAGAGCCAA---TCATGGGTGATGCCGATGAAATAGAGATA 219
QY 320 HisGlyTyrPheSerTyrGlyGlyTyrIleuSerLeuMetAlaLeuMetGlnArgSerAspIle 339
DB 220 CATGGCTGGAGCTATGATGCTTATGACTGAACTGATCTGATACGACGCGCATGATGC 279
QY 340 PheArgValAlaIleAlaGlyAlaIleProValThrLeuTrpIlePheTyrAspPheTrpIleYr 359
```

```

DB 280 TTCAAGTCGAGTAGCGGCGCTGTCTATGACTGGAATTCGATATGAGATTATGTAC 339
RESULT 10
US-09-389-681-428
; Sequence 428: Application US/09389681A
; Patent No. 6518237
; GENERAL INFORMATION:
; APPLICANT: Yugu, Jiang
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.470C3
; CURRENT APPLICATION NUMBER: US/09/389,681A
; CURRENT FILING DATE: 1999-09-02
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 428
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-389-681-428

Alignment Scores:
Pred. No.: 1,46-18 Length: 535
Score: 233.50 Matches: 52
Percent Similarity: 54.86% Conservative: 27
Best Local Similarity: 36.11% Mismatches: 54
Query Match: 12.08% Indels: 11
DB: Gaps: 3

US-10-070-464-7 (1-360) x US-09-389-681-428 (1-535)
QY 221 ProHisAspLeuGlnProGlyLysIleTyrProThrValLeuPheIleTyrGlyPro 240
DB 15 CCTCCTCAATTGACAGATCAAGAAAGATCCCTTGTAATTCAGATGATATGCTGCC 74
QY 241 GluValGlnLeuValAsnAsnArgPheLysGlyValTyrPheArgLeuAsnThrLeu 260
DB 75 TGCACTCAGAGTGTAAGTCT-----GTTTGGCTGTATATGGATA 116
QY 261 AlaSerLeu-----GlyTyrValValValIleAspAsnArgGlySerCys 276
DB 117 TCTTATCTTGCAGATGAGAAAGGATGCTCATTCCTTGTTGATGATGTCGAGAAACAGCT 176
QY 277 HisArgGlyLeuIysPheGluGlyAlaPheLysTyrIysMetGlyGlnIleGlnIleAsp 296
DB 177 TTCAGAGGTGACAAATCTCTCTATGCACTGATTCGAAAGCTGGGTATTATGAGATTGA 236
QY 296 AspGlnValGluGlyLeuGlnTyrIleuAlaSerArgTyrAspPheIleAspLeuAspArg 316
DB 237 GACCGATTTACAGCTGCTGAGAAATTCATA---GAAATGGCTTTCATTGATGAAAAAGA 293
QY 317 ValGlyIleHisGlyTyrPheSerTyrGlyGlyTyrIleuSerLeuMetAlaLeuMetGlnArg 336
DB 294 ATAGCCATATGAGGCTGCTCTATGAGAGATACGTTTCATCATCGCCCTTGCACTGGA 353
QY 337 SerAspIlePheArgValAlaIleAlaGlyAlaIleProValThrLeuTrpIlePheTyrAsp 356
DB 354 ACTGCTTTTCAATATGTGTATAGCAATGAGTCCAGCTCCAGCTGGAAATATTATACCGG 413
QY 357 ThrGlyTyrThr 360
DB 414 TCTGTCTACACA 425
RESULT 11
US-09-620-405B-428
; Sequence 428: Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
```



```
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/229,059
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/699,103
FILING DATE: 16-AUG-1996
APPLICATION NUMBER: 60/002,581
FILING DATE: 17-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31,259
REFERENCE/DOCKET NUMBER: 09272/005001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/322-5070
TELEFAX: 650/854-0875
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4982 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-229-059-1

Alignment Scores:
Pred. No.: 2.26e-14 Length: 4982
Score: 211.00 Matches: 49
Percent Similarity: 50.74% Conservative: 20
Best Local Similarity: 36.03% Mismatches: 55
Query Match: 10.92% Indels: 12
DB: 4 Gaps: 4

US-10-070-464-7 (1-360) x US-09-229-059-1 (1-4982)
QY 230 TyrProthrValleuPheileTYRGLYProGlnValGlnLeuValIasnAsnArgPhe 249
Db 4941 TATCCTGATTTTCTTTCATGATGAGGACCGAATTCACACAGTT-----4894
QY 250 LysGlyValLysTYRPhaArgLeuAsn-----ThrLeuAlaSerLeuGlyTyr 265
Db 4893 -----GTCAAAACGTTTCCGTGAGATTATGAGTGTAGCTTCACATTAAACGA 4840
QY 266 ValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGlnGlyAla 285
Db 4839 ATGTGATGTTGTGTGACGGTCTGTGACTGGCTTCAAGGTCAAGACTTATGATCCCTT 4780
QY 286 PheLysTYRlysmecGlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnTyrLeu 305
Db 4779 GTTCGCATAGGCTCGGTGATTAACAGGCCGCCGACCAAAATATCTGCGCTTCCTTAT 4720
QY 306 AlaSerArgTYRaspPheIleAspLeuAspArgValGlyIleHisGlyTTPSerTYRGLY 325
Db 4719 GGTTCCT--TTAACTTTGTGTGATCCGCAAAAGATTTCCTATTGTTGTGTGATACGG 4663
QY 326 GlyTYRleuSerleuMetAlaLeuMetGlnArgSerasp--IlePheArgValAlaIle 344
Db 4662 GGGTACCTGACACTTAACTTGTGAGAAAGATGCCGAGACACTTCAAAATACGGGATG 4603
QY 345 AlaGlyAlaProValThrLeuTrpIlePheTYRAspThrGlyTYRThr 360
Db 4602 TCACTTGCCGACGATACCGACTGAGATTATGATCTGTATTACT 4555
```

Search completed: October 16, 2003, 03:31:25
Job time : 81.4006 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 16, 2003, 00:05:32 ; Search time 269.152 Seconds

(without alignments)
3512.533 Million cell updates/sec

Title: US-10-070-464-7

Perfect score: 1993

Sequence: 1 EEDARSAGVATFVQDEPDR.....RVATAGAPVTIMTYDTCYT 360

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1750203 seqs, 1313063994 residues

Total number of hits satisfying chosen parameters: 3500406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODL=ftime+ p2n.model -DEV=x1h
-O=/cgn2_1/USPTO.spool/US10070464/runat_15102003_113555_24924/app_query.fasta_1.2652
-DB=Published Applications NA -OFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOECT=0 -LOOEXT=0 -UNITS=bits -STAR=1 -END=1 -MATRIX=blcsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEXT=pct -HEPESIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10070464 @CGN 1 1 845 @runatc_15102003_113555_24924
-NCPU=6 -ICPU=3 -NO MAP -LARGESUBSTR -NEG_SCORES=0 -WAIT -DSPLOK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCF_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCFUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	DB ID	Description
1	1880	97.3	4309	10 US-09-976-674-14 Sequence 14, Appl

2	1836.5	95.0	2649	12 US-10-054-776-1	Sequence 1, Appl
3	1836.5	95.0	2649	12 US-10-170-789-39	Sequence 39, Appl
4	1836.5	95.0	2671	10 US-09-976-674-2	Sequence 2, Appl
5	1836.5	95.0	3143	12 US-10-170-789-37	Sequence 37, Appl
6	1815.5	94.0	4829	10 US-09-976-674-12	Sequence 12, Appl
7	1522	78.7	4685	10 US-09-976-674-22	Sequence 22, Appl
8	1522.5	78.7	4676	10 US-09-976-674-20	Sequence 20, Appl
9	1229	63.6	4523	10 US-09-976-674-8	Sequence 8, Appl
10	1220.5	63.1	2617	10 US-09-976-674-4	Sequence 4, Appl
11	1220.5	63.1	4219	10 US-09-976-674-28	Sequence 28, Appl
12	1220.5	63.1	4302	10 US-09-976-674-24	Sequence 24, Appl
13	1201	62.1	4180	10 US-09-976-674-36	Sequence 36, Appl
14	1183.5	61.2	4263	10 US-09-976-674-34	Sequence 34, Appl
15	1183.5	61.2	3262	13 US-10-098-841-83	Sequence 83, Appl
16	1169.5	60.5	2801	13 US-10-098-841-100	Sequence 100, Appl
17	1132.5	58.6	4076	10 US-09-976-674-32	Sequence 32, Appl
18	1132.5	58.6	4159	10 US-09-976-674-40	Sequence 40, Appl
19	1113	57.6	4037	10 US-09-976-674-30	Sequence 30, Appl
20	1113	57.6	4120	10 US-09-976-674-38	Sequence 38, Appl
21	630	32.6	502	11 US-09-918-995-19585	Sequence 19585, A
22	368	19.0	561	11 US-09-764-891-877	Sequence 877, App
23	352	18.2	2079	13 US-10-044-090-843	Sequence 843, App
24	352	18.2	2411	10 US-09-976-674-26	Sequence 26, Appl
25	315	16.3	3407	12 US-10-423-714-5	Sequence 5, Appl
26	315	16.3	3407	13 US-10-002-593-5	Sequence 5, Appl
27	315	16.3	3407	14 US-10-165-603-6	Sequence 6, Appl
28	308.5	16.0	2130	10 US-10-156-761-131	Sequence 3131, Ap
29	308.5	16.0	9025608	14 US-10-156-761-1	Sequence 1, Ap
30	297.5	15.4	4835	10 US-09-917-8004-1570	Sequence 1570, Ap
31	297.5	15.4	4835	10 US-10-165-603-5	Sequence 5, Appl
32	271	14.0	4797	11 US-09-764-891-7074	Sequence 7074, Ap
33	261	13.5	2812	14 US-10-084-817-24	Sequence 24, Appl
34	258.5	13.4	2366	12 US-10-101-510-683	Sequence 683, Appl
35	258.5	13.4	2788	12 US-10-269-909-22	Sequence 22, Appl
36	258.5	13.4	3138	14 US-10-198-846-13171	Sequence 13171, A
37	257.5	13.3	3224	12 US-10-240-965-117	Sequence 117, App
38	257	13.3	2814	10 US-09-962-832-108	Sequence 108, App
39	257	13.3	2814	10 US-09-954-456-1148	Sequence 1148, Ap
40	257	13.3	2814	10 US-09-954-531-367	Sequence 367, App
41	257	13.3	2814	12 US-10-101-510-30	Sequence 30, Appl
42	257	13.3	2814	12 US-10-301-822-54	Sequence 54, Appl
43	257	13.3	2814	12 US-09-873-367C-499	Sequence 499, App
44	257	13.3	2814	14 US-10-177-293-135	Sequence 135, App
45	257	13.3	2815	9 US-09-265-606-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-976-674-14
; Sequence 14, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Julien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 4309
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-14

Alignment Scores: 5.23e-231 Length: 4309
Pred. No.: 1

Score: 1880.00 Matches: 358
Percent Similarity: 98.90% Conservative: 0
Best Local Similarity: 98.90% Mismatches: 2
Query Match: 97.26% Indels: 3
DB: 10 Gaps: 0
US-10-070-464-7 (1-360) x US-09-976-674-14 (1-4309)

QY 1 GIUUIAASPALAARSGERAGLYVALAATHRPHEVALLEUNGILUGIUPHEASPAR 20
DB 943 GAAGAGAGATGCCAGATCAAGCTGAGTCCGCTACCTTGTCTCCAGAGAAATTTATAGA 1002
QY 21 TYRSEGIYTYRTRIPCYSPROLYSAGIUTRTHPROSEGIYGLYLSILEU 40
DB 1003 TATCTGGCTATTGGTGGTGTCCAAAGCTGAAACACTCCAGAGGGTGGTAAATTCCT 1062
QY 41 ARGILEUITYRGILUGIUNASAPGIUSERGIVAGIUILEIHSIVALTHSEPRO 60
DB 1063 AGAATTCATATAGAGAAATGATGATCTGAGTGGAATTTATTCATCTTACATCCCT 1122
QY 61 METLEUGIUTHRARGARALASPSERPHEARGTYRPROLYETHRGITYTHRALASPRO 80
DB 1123 ATGTTGGAAACAAGAGGCGACATTCATTCGTTATCCTAACAAGTACAGCAATCCT 1182
QY 81 LYSVALTHRPHELYMESCERGIUILEMETLEASPALAGIUGIYARGILEILEVALASP 100
DB 1183 AAAGTCACCTTTAAAGATGCAGAAATTAATGATTGATGCGAAGGAATCCAA-GTTGAT 1241
QY 101 GIUVALARGARLEUVALTYRPHEGIUGITYRLYSASPSPERPROLEUGIUIHSILEU 120
DB 1242 GAAGTCAGAGAGCGGTATATTTTGAAGGACCAAGACATCCCTTAGAGCATCACCTG 1301
QY 121 TYRVALIASERTYRVALASPROGIYGVALLTRARGLEUTHRASPARGLTYRYSER 140
DB 1302 TAGGATGATAGTAGTAATCTCGAGAGTGACAGGCTACGACCGTGCTACTCA 1361
QY 141 HISSECYCYAIISESGINHSICYSASPHEPHEIISESLYSESYSEZASNGIUNYS 160
DB 1362 CATCTTGCTGATCAGTCAGACATGTCATCTTTATTAAGATAGATAGTAAACAGAA 1421
QY 161 AANPROHSICYSVALSERLEUTYRLYSLEUSERSESPROGIUNASPARPROTHRCYLS 180
DB 1422 AATCCACACTGTGTGCCCTTTACAGCTATCAAGCTCCGGAAGATGACCCCACTTGCAA 1481
QY 181 THRIVSGIUPHEITRALATHRIILEUNASPSERAGLYPROLEUPROKSPYRTHRPRO 200
DB 1482 ACAAGAGAAATTTGGCCACATTTTGAFTCAAGAGTCTCTTCGACTACTACTCT 1541
QY 201 PROGIUILEPHESESPHEGIUSERTHTRGIYPHEITHLEUTYRGIYMETLEUTYRYS 220
DB 1542 CCAGAAATTTCTCTTTGAAAGTACTACTGATTTACATGTGATGGGATGCTCTACAA 1601
QY 221 PROHSIASPHEUGIUPROGIYLYSISTYRPROTHRVALLEUPHEIIERYTGLYLY-P 240
DB 1602 CCTCATGATCTACAGCTGGAAGAAATATCTACTAGCTGTCTCATATATATGTTGCTC 1661
QY 240 ROGIVAGIUNLEUVALAENASNAKSPHELYSGLYVALIYSTYRPHEARGLAENATHRL 260
DB 1662 CTCAGGTGCAATGGTGGATTAATCGSTTTAAAGGATCAAGATTTCCGCTGAATACC 1721
QY 260 EUALISERIEUGIYTYRVALIIVALIIEASPSANARGIYSECYSHISARGIYLY 280
DB 1722 TAGCCTCTCTAGCTTATGTGTTGTAGTATAGACAACAGGGGATCTCTGACCAAGGC 1781
QY 280 EUIYEPHEGIUNGLIYLAPELYSTYRLYSMERGILGNIILEGIUILEASPARPGIINVAIG 300
DB 1782 TTAAATTTTGAAGCGCTTTAAATATATAATAGAGATGCGATAGAGGTG 1841
QY 300 IUGIYLEUGINTYRLEUALASERARGTYRASPHEIIASPLEUNASPARVALIYIIEH 320
DB 1842 AAGGACTCCAATATCTAGCTTCTCGATATGATTTGACTTAGATGCTGAGGCGATCC 1901
QY 320 ISGLYTRPSERTYRGLYIYTYRLEUSERLEUMETALALEUMETGINATGSEASPIIEP 340

DB 1902 ACCGCTGATCTTATGAGATACCTCTCCGATGATGATATGACAGGTAGATATCT 1961
QY 340 HEARGVALAIIIEALAGIYALAPROVALTHRLEUTRIPHEIETHYRASPHTRGITYRT 360
DB 1962 TCAGGGTGTATTCTCGGGGCCAGTCACTGTGTGATCTTATGATACAGATATCA 2021
QY 360 hr 360
DB 2022 CG 2023

RESULT 2

US-10-054-776-1
; Sequence 1, Application US/10054776
; Publication No. US20030165818A1
; GENERAL INFORMATION:
; APPLICANT: Mark Robert Edbrooke
; APPLICANT: Alan Peter Lewis
; TITLE OF INVENTION: NOVEL PROTEIN
; FILE REFERENCE: OG1042US
; CURRENT APPLICATION NUMBER: US/10/054,776
; CURRENT FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2649)
US-10-054-776-1

Alignment Scores:

Pred. No.: 9,81e-226 Length: 2649
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 12 Gaps: 1

US-10-070-464-7 (1-360) x US-10-054-776-1 (1-2649)

QY 1 GIUUIAASPALAARSGERAGLYVALAATHRPHEVALLEUNGILUGIUPHEASPAR 20
DB 730 GAAGAGAGATGCCAGATCAAGCTGAGTCCGCTACCTTGTCTCCAGAGAAATTTATAGA 789
QY 21 TYRSEGIYTYRTRIPCYSPROLYSAGIUTRTHPROSEGIYGLYLSILEU 40
DB 790 TATCTGGCTATTGGTGGTGTCCAAAGCTGAAACAACTCCAGTGTGTAAATTCCT 849
QY 41 ARGILEUITYRGILUGIUNASAPGIUSERGIVAGIUILEIHSIVALTHSEPRO 60
DB 850 AGAATTCATATAGAGAAATGATGATCAATCTGAGGAGAAATTTATCATGTTACATCCCT 909
QY 61 METLEUGIUTHRARGARALASPSERPHEARGTYRPROLYETHRGITYTHRALASPRO 80
DB 910 ATGTTGGAAACAAGAGGCGACATTCATTCGTTATCCTAACAAGTACAGCAATCCT 969
QY 81 LYSVALTHRPHELYMESCERGIUILEMETLEASPALAGIUGIYARGILEILE- 98
DB 970 AAAGTCACCTTTAAAGATGCAGAAATTAATGATGCTGGAAGAGATCATATGATGTC 1029
QY 98 ----- 98
DB 1030 ATAGATTAAGAACTATATTCACCTTTTAGATTTATTTGAAGAGTTGAATATATGTC 1089
QY 98 ----- 98
DB 1090 AAGCTGATGGAATCTCGAGGAAATATATGCTGTTCATCTACTAGATGCTCCAG 1149
QY 98 ----- 98

Db 1150 ACTGCCTGAGATAGTGTGATCTCACCTGAATTATTTATCCAGTAGAAGATGATGTT 1209
QY 98 ----- 98
Db 1210 ATGGAAGGAGAGACTGATTAAGTCAAGTCCGATTCGTGACGCCACTAATTATCTAT 1269
QY 98 ----- 98
Db 1270 GAAGAAACACAGACATCTGGATTAATATCCATGACATCTTTGATGTTTTTCCCAAGT 1329
QY 98 ----- 98
Db 1330 CACGAGAGGAATTGAGTTTATTTTGGCTCTGATGCAAAACAGGTTCCGTCATTTA 1389
QY 98 ----- 98
Db 1390 TACAAATTTACATCTATTTTAAAGGAAACAAATATAACATCCAGTGTGGCTGCT 1449
QY 98 ----- 98
Db 1450 GCTCCAAATGATTTCAAGTGTCTTATCAAAAGAGGATAGCAATTACAGTGTGAATGG 1509
QY 99 -----ValAspGluValArgLeuValTyr 1510
Db 1510 GAAGTTCTGGCCGCGCATGATCTAATATCCAAAGTTGATGAAGTCAGAAAGCTGGTATAT 1569
QY 108 PheGluGlyThrIleAspSerProLeuGluHisIleuTyrValValSerTyrValAsn 1629
Db 1570 TTTGAAAGGACCAAAAGACTCCCTTTAAGCATCACTGGTACGTGATGATTAAT 1629
QY 128 ProGluIleuValThrArgLeuThrAspArgIleSerHisSerCysCysIleSerGln 1630
Db 1630 CCTGAGAGGTGACAAAGGCTGATGCTGACCGTGGCTGCTGATCTGCTGATCAGTCAAG 1689
QY 148 HisCysAspPhePheIleSerIleTyrSerAsnGlnYasnProHisCysValSerLeu 1690
Db 1690 CACGTGACTTCTTATAGTATAGTATAGTAAACGAAAGATCCACACGTGTGTCCCTT 1749
QY 168 TyrIleuSerSerProGluAspAspProThrCysIleThrIleGluPheTrpAlaThr 187
Db 1750 TACAAGCTATCAAGCTCCTGAAGATGACCACCTGCAAAACAAAGAAATTTTGGGCCACC 1809
QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 207
Db 1810 ATTTTGATTCAGCAGGTCTCTCTGATCTATCTCTCCGAAATTTTCTCTTTGAA 1869
QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleProHisAspLeuGlnProGly 227
Db 1870 AGTACTACTGGATTTACATTTGATGATGATGCTCTCAAGCCCTCATGATCTACAGCTGGA 1929
QY 228 LysIleTyrProThrValLeuPheIleTyrGlyIleProGluIleGluLeuValAsnAsn 247
Db 1930 AAGAAATATCTACTACTGCTCTCTCATATATGCTGCTCAGGTGCAGTGTGATATAT 1989
QY 248 ArgPheIleGlyValIleTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db 1990 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTCTGAGTTATGTGTT 2049
QY 268 ValValIleAspAsnArgIleYserCysHisArgGlyLeuIlePheGluGlyValAlaPheIle 287
Db 2050 GTAGGATAGACAAACAGGGGATCCCTGTCAACGAGGGCTTAAATTTGAAGGCGCTTTAAA 2109
QY 288 TyrIleMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db 2110 TATATAAAGGATGCAATGAAATGACATCGATCGGGAAGAGCTCAATATCTTACCTTCT 2169
QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyIleTyr 327
Db 2170 CGATATGATTTTCATTTGATCTTGAATGTGTGGGCACTCAAGCTGTCTCTATGAGAGATAC 2229
QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db 2230 CTCCTCCGATGAGCTATATGACAGGTACAGTATCTTCAGGGTGTGCTATGTGCGGCC 2289

QY 348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
Db 2290 CCAGTCACTGTGTGATCTCTATGATATACAGATACAG 2328
RESULT 3
US-10-170-789-39
Sequence 39, Application US/10170789
Publication No. US20030180930A1
GENERAL INFORMATION:
APPLICANT: Rachel E. Meyers
APPLICANT: Olandt, Peter J.
APPLICANT: Kapeller-Libermann, Rosana
APPLICANT: Curtis, Rory A. J.
APPLICANT: Williams, Mark
APPLICANT: Weich, Nadine
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
FILE REFERENCE: 10448-191001
CURRENT APPLICATION NUMBER: US/10/170,789
CURRENT FILING DATE: 2002-06-13
PRIOR APPLICATION NUMBER: US 09/797,039
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: PCT/US01/06525
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/186,061
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 09/882,166
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: PCT/US01/19269
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/212,078
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 09/934,406
PRIOR FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: PCT/US01/26052
PRIOR FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: US 60/226,740
PRIOR FILING DATE: 2000-08-21
PRIOR APPLICATION NUMBER: US 09/861,801
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: PCT/US01/16549
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: US 60/205,508
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: US 09/801,267
PRIOR FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: PCT/US01/07138
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: US 60/187,454
PRIOR FILING DATE: 2000-03-07
PRIOR APPLICATION NUMBER: US 09/829,671
PRIOR FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: PCT/US01/40483
PRIOR FILING DATE: 2001-04-11
PRIOR APPLICATION NUMBER: US 60/197,508
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 09/961,721
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: PCT/US01/29904
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 60/235,023
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US 10/045,367
PRIOR FILING DATE: 2001-11-07
PRIOR APPLICATION NUMBER: US 60/246,561
PRIOR FILING DATE: 2000-11-07
PRIOR APPLICATION NUMBER: US 09/801,275
PRIOR FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: PCT/US01/07074
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: US 60/187,420
PRIOR FILING DATE: 2000-03-07

```

; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-170-789-39

Alignment Scores:
Pred. No.:          9,81e-226      Length:      2649
Score:             1836.50         Matches:     360
Percent Similarity: 67.54%         Conservative: 0
Best Local Similarity: 67.54%      Mismatches:  0
Query Match:       95.01%          Indels:      173
DB:                12              Gaps:        1

US-10-070-464-7 (1-360) X US-10-170-789-39 (1-2649)

QY      1 GIUGLUAAPALARGSERIAGLYVALATHPHEVAILEUGLNGLUHPHEASPAR 20
DB      730 GAGAAAGATCCAGATCAGCTGAGAGCTGACCTTTGTTCTCCAGAAAGAAATTTGATGA 789
QY      21 TYRSEGLTYRTPTRPCYSPROLYSALGLUTHPROSEGLYGLYLSILEU 40
DB      730 TATCTGCTATTGGTGGTGTCCAAAGCTGAACCACTCCAGTGTGTAAATTTCTT 849
QY      41 ARGILEUITYRGLUGLUNASPGIUSERGLUVALGLILEHISVALTHRSERPRO 60
DB      850 AAAAAATCTATATGAAAGAAATGATGATCTGAGGTGGAATTAATTAATTAATTCATCCCT 909
QY      61 METLEUGLUTHRARGRGLAASPSERPHARGTYRPROLYSTHRGLYTHRALAASPRO 80
DB      910 ATCTTGAAACAAGAGGCGCAGATTCATTCCTGTAATCCAAAACAGGTACAGCAATCTT 969
QY      81 LYSVALTHRPHLYSMETSERGLULEMETILEASPALEGLUGLYARGILELE----- 98
DB      970 AAAAGTCACTTTTAAGATGTCAGAAATAATGATGATGCTGAAGAGATCATAGATGTC 1029
QY      98 ----- 98
DB      1030 ATATGATAGGAATTAATTAACCTTTTGAGATTCATTTGAAAGAGTTGAATTAATTCGCC 1089
QY      98 ----- 98
DB      1090 AAGAGTGATGAGCTCTGAGGGAATAATGCTTGTCATCTACTAGATGCTCCAG 1149
QY      98 ----- 98
DB      1150 ACTCGCTGCAATAGTGTGATCTGCACCTGAATTAATTAATCCAGTAGAAGATGATGTT 1209
QY      98 ----- 98
DB      1210 ATGGAAAGCAGAGACTCATTTGAGTCAGTGCCTGATTCGTGAGGCCATAATTAATCTAT 1269
QY      98 ----- 98
DB      1270 GAAAGAAACAACAGACATCTGATTAATATCATGACATCTTTTCATGTTTTCGCCAAGT 1329
QY      98 ----- 98
DB      1330 CACGAAAGAGAAATGAGTTTATTTTGGCTCTGAATGCAAAAACAGGTTTCGTCATTTA 1389
QY      98 ----- 98
DB      1390 TACAAAATTACATCTATTTTAAAGAAAGCAATATAAGATCCAGTGTGGGCTGCCCT 1449
QY      98 ----- 98
DB      1450 GCTCAAGTAGATTCAAGTCTCTATCAAGAGAGATGCAATTACCACTGGTGAATG 1509
QY      99 ----- ValASPLUVALARGHLEUVALTYR 107
DB      1510 GAAGTTCTTGCGCGCATGATCTAATATCCAAAGTTGATGAGCTGAAAGCTGGTATAT 1569
```

```

QY      108 PHEGLUGLYTHRNYLASEPSERPROLEUGLUNHSHILEUITYRVALSERITYRVALASN 127
DB      1570 TTGAAAGCACCAAAAGACTCCCTTTAAGACATCCCTGTAGTGTAGTACGTAAT 1629
QY      128 PROGLYGLUVALTHRARGLEUTHRASPARGLYTYRSETHLSERHCYSCYSILESERGIN 147
DB      1630 CCTGAGAGGTGACAGAGGCTGACTGACCGTGGCTACTACATTTCTGTCATCATGTCAG 1689
QY      148 HISCYASPPHEPHEILESERLYSTYRSEANGLYASNPROMHISCYSVALSERLEU 167
DB      1630 CACTGTGACTCTTTATTAAGTATGTAATACCAAGAAATCCACACAGTGTGTGCCCTT 1749
QY      168 TYRYSLEUSERSERPROGLUNASPSERPROTHRCYSLYETHLYGLUHPHETPLAATHR 187
DB      1750 TACAAAGCTATCAAGTCTGAGAGATACCCACTTGCAAAAACAAAGAAATTTGGGCACCC 1809
QY      188 ILEUAASPSERIALGLYPROLEUPROASPTYRTHRPROPROGLUULEPHESERPHEGLU 207
DB      1810 ATTTGGATTCAGCAGAGTCTCTCTGACTATACCTCCAGAAATTTCTCTTTGAA 1869
QY      208 SEPTHRTNGLYPHETHLEUTHYRGLYMETLEUTHYRYSPROHISASPLEUINPROGLY 227
DB      1870 AGTACTAGTGAATTAATGATGATGATGCTCTACAAACCTCATGATCATACAGCCGTGA 1929
QY      228 LYSLYSTYRPROTHRVALLEUPHEILETYRGLYGLYPROGLYVALGLNLEUVALASNASN 247
DB      1930 AAAAAATCTCTACTGCTGTGCTGTTCAATATATGATGATGCTCTCAAGTGCAGTTGGTAATAT 1989
QY      246 ARGPHLYSGLYVALLYSTYRPHEARGLUEANTRHLEUALASERLEUGLYTYRVAL 267
DB      1990 CGGTTTAAAGAGTCAAGATTTCCGCTTGAAATCCCTGATCTCTAGATTATGATG 2049
QY      268 VALVALILEASPSANRGLYSERCYSHISARGLYLEULYEPHEGLUGLYALAPHELYS 287
DB      2050 GTAGGATGTGACAAAGGCGATCTGTCCACCAAGGCTTAAATTTGAAGCGGCTTTAAA 2109
QY      288 TYRYSMETGLYGLNILEGLULILEASPSAGLUNVALGLUGLYLEUHTYRLEUALASER 307
DB      2110 TATTAATAAGGTGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2169
QY      308 ARGTYRASPHEILEASPLEUSPARGLYGLYILEHISGLYTRPSERTYRGLYGLYTYR 327
DB      2170 CGATATGATTTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2229
QY      328 LEUSERLEUMETALALEUMETGLNARGSERASPILEPHARGVALAALILEAGLYALA 347
DB      2230 CTCTCCCTGATGCAATTAATGACAGGTGAGATATCTTCAAGGTTGCTATTGCTGGGCC 2289
QY      346 PROVALTHRLEUTHRIEPHETYSRPTHRGLYTYRTHR 360
DB      2290 CAGTCACTCTGTGATCTTCTATGATACAGATACAGC 2328

RESULT 4
; Sequence 2, Application US/0976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Junien, Jean-Louis
; APPLICANT: Riviere, Pierre
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2671
; TYPE: DNA
```

ORGANISM: Homo sapiens
US-09-976-674-2

Alignment Scores:

Pred. No.: 9,946-226 Length: 2671
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 10 Gaps: 1

US-10-070-464-7 (1-360) x US-09-976-674-2 (1-2671)

```

QY      1  GUGUASPDAAAGSERRAAGLYVALATHRPHEVALLEUGLUGLUGLUPHESAPRG 20
DB      737  GAAGAAGATGCCAGATCGAGCTGGAGTCGCTACCTTGCTTCACAGAAAGATTGATAGA 796
QY      21  TYSERGLYTRTTRPCYSPROLYSALAGIUTRTHPRROSERGLVGLVYSLELEU 40
DB      797  TATTCGTGCTATTGGTGGTGTCCAAAGCTGAACACTCCAGTGGTGTAATAATCTT 856
QY      41  ARGILEUITYRGLUGLUASNAAPGLUSERGLUVALGULLELEHISVALTHRSEPRO 60
DB      857  AGAATTCATATGAGAAATATGATGATCTGAGCGAAATTATTCATGTTACATCCCT 916
QY      61  METLEUGLUTRARGARGALASPSEPRHEATGLYRPROLYSTHGLYTHRALASNPRO 80
DB      917  ATGTGGAAACAGAGAGGCGACATTCATTCCTTATCCTAACACAGTACACGCAATCCT 976
QY      81  LYSVALTHRPHELYMETSERGLULEMETILEASPDALGULGARIGLELE----- 98
DB      977  AAAAGTCACTTTTAAAGATGTCAGAAATTAATGATTGATGCTGAAGAGATCATAGATGTC 1036
QY      98  ----- 98
DB      1037  ATAGATAAGAACTAATTCACCTTTTGAGATTCTATTGAAGAGATTGAATATATTGCC 1096
QY      98  ----- 98
DB      1097  AGAGCTGATGAGCTCTGAGGAGAAATATGCTTGTCATCTTAAGATGCTCCAG 1156
QY      98  ----- 98
DB      1157  ACTGCCTGCAGATAGTGTTCATCTCACCTGAATTAATTATCCACAGTAGAAGATGTT 1216
QY      98  ----- 98
DB      1217  ATGGAAGGACAGACTCATTTGATGATGAGTCTGATCTGTGACGCCACTAATATCTAT 1276
QY      98  ----- 98
DB      1277  GAAGAAACACAGACATCTGATTAATAATCCATGACATCTTTCATGTTTTCCCAAGT 1336
QY      98  ----- 98
DB      1337  CACGAAGAGAAATTGAGTTATTTTGGCTCTGAATGCAAAACAGTTCCGTCAATTTA 1396
QY      98  ----- 98
DB      1397  TACAAATATACATCTATTTTAAAGAAAGCAATATAACATCCAGTGGTGCGCT 1456
QY      98  ----- 98
DB      1457  GCTCCAGATGATTCAAGTGTCTTATCAAGAAGAGATGCAATTACAGTGTGAATGG 1516
QY      99  -----ValaspGLuValArgArgLeuValTyr 107
DB      1517  GAAGTCTTGGCCGCGCATGATCTAATATCCAAAGTTGATGAAGTCAAGAGCGCTGATAT 1576
QY      108  PHEGLUGLYTHRLYASPSERPROLEUGLUNISHISLEUITYRVALVASERTYRVALSN 127
DB      1577  TTTGAAGGACCAAGAGACTCCCTTTAGAGCATCACCTGTAGTGTAGTACGTACGTAAAT 1636

```

```

QY      128  PROGLUGLUVALTHRARGLEUTHRASPARGLYTRSERHISSECYCSALLSEGIN 147
DB      1637  CCTGGAGGAGTGACAAAGGCTGACGACCGTGCTACTACATCTTTCGTCAGTCAG 1696
QY      148  HISCYASPPHEPHEILLESTYSTRSERNSGLUNLVSANPROHISCYVALSERLEU 167
DB      1697  CACTGTACTCTTTATATAGTATAGTATACCGAAGAAATCCACACTGTGTCTCTT 1756
QY      168  TYRYSLEUSERSERPROGLUASPARPROTHRCYSLYSTRHLYEGLUPHETPALATHR 187
DB      1757  TACAAAGTATCAAGTCCGGAAGATGACCCCAACTCCAAACAAAGAAATTTTGGCCACC 1816
QY      188  ILEUASPSERIALAGLYPROLEUPROASPTYRTHRPROFROGLUNILEHESERPHES 207
DB      1817  ATTTGGATTCAGCAGCGCTCTTCTTCTACATACCTCCCGAAATTTCTCTTTGAA 1876
QY      208  SESTRHTRHGLYPHETHRLEUTHYRGLYMETLEUITYRVSPROHISASPLENGINPRO 227
DB      1877  AGTACTACTGGATTTACATTTGATATGAGATGCTCTACAGCTCATGATCTACAGCTGGA 1936
QY      228  LYSLYSTRPROTHRVALLEUPHEILLETYRGLYGLYPROGLUNVALGULEUVALASNA 247
DB      1937  AAGAAATATCTCACTGCTGCTGTTATATGTTGCTCCAGTGCAGTTGTTGATATAT 1996
QY      268  ARGPHELYSGLYVALLYSTRPHEARGLEUENHTRLEUALASERLEUGLYTYRVALVAL 267
DB      1997  CGATTTAAAGAGTCAAGTATATTCGCTTGAATACCTTAAGCTCTCTAGATTATGGTGT 2056
QY      268  VALVALILEASPARARGLYSERCYSHISARGGLYLEUVPHEGLUGLYLAAPHELYS 287
DB      2057  GTAGTGATGACACAGAGGAGATCTGTACCCAGAGGCTTTAAATTTGAAGGCCCTTTAA 2116
QY      288  TYRYSMETGLYGLNILEGLNILEASPARGLUVALGULYLEUGLINTYRLEUALASER 307
DB      2117  TATAAATAGGGTCAATATGAATATGACGATCGAGGAAAGACTCCAAATATCTAGCTTCT 2176
QY      308  ARGTYRASPHEILLEASPLEUSPARGLYGLYILEHISGLYTRPSERTYRGLYGLYTYR 327
DB      2177  CGATATGATTCATGACTGTAGTATGCTGTGGCATCCAGCGCTGTGCTTATGAGACATAC 2236
QY      328  LEUSERLEUMETALILEUMETGLNARGSERASPIIPEHARGVALALILEALAGLYALA 347
DB      2237  CTCTCCCTGATGGCATTTATGAGAGGTCAGATATCTTCAAGGATGTCTATTTCTGGGCCC 2296
QY      348  PROVALTHRLEUTHRILEPHEITYRASPETHRYTYRTHR 360
DB      2297  CCAATGCTCTGTGATCTTCTATGATACAGATATCAACG 2335

```

RESULT 5
US-10-170-789-37
Sequence 37, Application US/10170789
Publication No. US20030180930A1
GENERAL INFORMATION:
APPLICANT: Rachel E. Meyers
APPLICANT: Olandt, Peter J.
APPLICANT: Kapeller-Libermann, Rosana
APPLICANT: Curtis, Rory A. J.
APPLICANT: Williamson, Mark
APPLICANT: Welch, Nadine
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
FILE OF INVENTION: AND PROTEASE FAMILY MEMBERS AND USES THEREOF
FILE REFERENCE: 10448-191001
CURRENT APPLICATION NUMBER: US/10/170,789
CURRENT FILING DATE: 2002-06-13
PRIOR APPLICATION NUMBER: US 09/797,039
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: PCT/US01/06525
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/186,061
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 09/882,166
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: PCT/US01/19269

```

/ PRIOR FILING DATE: 2001-06-15
/ PRIOR APPLICATION NUMBER: US 60/212,078
/ PRIOR FILING DATE: 2000-06-15
/ PRIOR APPLICATION NUMBER: US 09/934,406
/ PRIOR FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: PCT/US01/26052
/ PRIOR FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: US 60/226,740
/ PRIOR FILING DATE: 2000-08-21
/ PRIOR APPLICATION NUMBER: US 09/861,801
/ PRIOR FILING DATE: 2001-05-21
/ PRIOR APPLICATION NUMBER: PCT/US01/16549
/ PRIOR FILING DATE: 2001-05-21
/ PRIOR APPLICATION NUMBER: US 60/205,508
/ PRIOR FILING DATE: 2000-05-19
/ PRIOR APPLICATION NUMBER: US 09/801,267
/ PRIOR FILING DATE: 2001-03-06
/ PRIOR APPLICATION NUMBER: PCT/US01/07138
/ PRIOR FILING DATE: 2001-03-05
/ PRIOR APPLICATION NUMBER: US 60/187,454
/ PRIOR FILING DATE: 2000-03-07
/ PRIOR APPLICATION NUMBER: US 09/829,671
/ PRIOR FILING DATE: 2001-04-10
/ PRIOR APPLICATION NUMBER: PCT/US01/40483
/ PRIOR FILING DATE: 2001-04-11
/ PRIOR APPLICATION NUMBER: US 60/197,508
/ PRIOR FILING DATE: 2000-04-18
/ PRIOR APPLICATION NUMBER: US 09/961,721
/ PRIOR FILING DATE: 2001-09-24
/ PRIOR APPLICATION NUMBER: PCT/US01/29904
/ PRIOR FILING DATE: 2001-09-24
/ PRIOR APPLICATION NUMBER: US 60/235,023
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: US 10/045,367
/ PRIOR FILING DATE: 2001-11-07
/ PRIOR APPLICATION NUMBER: US 60/246,561
/ PRIOR FILING DATE: 2000-11-07
/ PRIOR APPLICATION NUMBER: US 09/801,275
/ PRIOR FILING DATE: 2001-03-06
/ PRIOR APPLICATION NUMBER: PCT/US01/07074
/ PRIOR FILING DATE: 2001-03-05
/ PRIOR APPLICATION NUMBER: US 60/187,420
/ PRIOR FILING DATE: 2000-03-07
/ NUMBER OF SEQ ID NOS: 63
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 37
/ LENGTH: 3143
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (229)...(2874)
US-10-170-789-37

```

Alignment Scores:

```

Pred. No.: 1,29e-225 Length: 3143
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 12 Gaps: 1

```

US-10-070-464-7 (1-360) x US-10-170-789-37 (1-3143)

```

QY 1 GIUGLHASPAAARGSERLAGLYVALAIAATHPhValleuGluGluGluPheaspArg 20
DB 958 GAAGAAGATGCCGATCAGCTGAGTCGCTAACCTTGTCTCCAAAGAATTGATGAGA 1017
QY 21 TYRSEGLTYRTPTIPCYSPROLYSALAAGLUTHThProserGlyGlyLysileu 40
DB 1018 TATTCGGCTATTGGTGGTGTCCAAAAGCTGAAACAACCTCCAGTGTGTAATAATCTT 1077
QY 41 ARGILEUITYRGILUGLHASPAGLUSERGLUVALGHLLEIHISVALThrSerPro 60

```

```

DB 1078 AGAATTCATATGAAGAAAATGATGAATCTGAGTGGAAAATATTCATGTTACATCCCT 1137
QY 61 MetleuGluThrArgArgAlaaspSerPheargIYrProLYSThrGlyThraIaaspPro 80
DB 1138 ATGTTGAAACAAAGAGGCGAGATTCATTCGTTATCTCTAAACAGGTACAGCAATCTCT 1197
QY 81 LysValIThrPheIysMetSerGluIleMetIleaspAlaGluIArgIleIle----- 98
DB 1198 AAAAGTCACTTTTAAATGATGCAAAATAATGATGATGCTGAAAGAAAGATCATAGATGTC 1257
QY 98 ----- 98
DB 1258 ATAGATAAGAACTAATTCACCTTTTGAGATTCATTGAAAGAGTTGAATATATATGCG 1317
QY 98 ----- 98
DB 1318 AGAGCTGATGGAATCTCTGAGGAAAATATGCTGTGTCATCTACTAGATGCTCCAG 1377
QY 98 ----- 98
DB 1378 ACTGCGCTGCAGATAGTGTGATCTCACCTGAATTAATTCACAGTAGAAGATGTT 1437
QY 98 ----- 98
DB 1438 ATGAAAAGCAGAGACTCATTTGAGTCAGTCAATTCGTGACCCCACTAATATCTAT 1497
QY 98 ----- 98
DB 1498 GAAGAAACAAACAGACATCTGATAAATATCCATGACATCTTCATGTTTCCCAAGT 1557
QY 98 ----- 98
DB 1558 CACGAAGAGAAATGAGTTTATTTTGGCTCTGAATGCAAAACAGTTCCGTCAATTA 1617
QY 98 ----- 98
DB 1618 TACAAATTAATCATCTATTTTAAAGAAACAAATTAACATCCAGTGGTGGCTGCC 1677
QY 98 ----- 98
DB 1678 GCTCCAGATGATTTCAAGTGTCTTATCAAGAAGAGATGCAATTACAGTGTGATG 1737
QY 99 ----- 107
DB 1738 GAAGTTCTTGCGCGGATGATCTAATATCAAGTTGATGAGTCAAGAGCTGATATAT 1797
QY 108 PheGluGlyThrIysaspSerProleuGluIshIseuTYrValIserTYrValasn 127
DB 1798 TTTGAAGCACCAGAGACTCCCTTTAGAGCATCATCTGATAGTACGTTACGTAAT 1857
QY 128 ProGlyGluValIThrArgLeuThrAspArgIYrSerHisSerCysSylIleSerGln 147
DB 1858 CCTGAGAGGTGACAAAGCTGACGACCTGCTACTCACTTCTTGTCGACACATCAG 1917
QY 148 HisCysaspPhePheIleSerIysTYrSerasnGluIysAspProHisCysValIserLeu 167
DB 1918 CACTGTGACTTTCTTTATAGTATGATATATACCAAGAAATCCACACTGTGTCTCT 1977
QY 168 TYrLYSLeuSerSerProGluaspAspProThrCysLYSThrLYSGluPheIValThr 187
DB 1978 TACAAGCTATCAAGTCTGAAATGACCACTGCGAAAACAAAGAAATTTTGGCCACC 2037
QY 188 ILeuaspSerIArgIYrProleuProaspTYrThrProProGluIlePheSerPheGlu 207
DB 2038 ATTTTGATTCAGCAGAGTCTCTCTGACTAATACCTCCAGAAATTTCTTTTGA 2097
QY 208 SerThrThrGlyPheThrIleuTYrGlyMetIleuTYrLYSProHisaspIleuGlnProGly 227
DB 2098 AGTACACTGAGATTTATGATGAGAGCTCTAACAACCTCATATATCTACAGCTGGA 2157
QY 228 LysLYSTYrProThrValIleuPheIleTYrGlyGlyProGluValGluIleuValasn 247

```

```

Db      2158 AAGAAATATCCACTGTGCTGTTCAATATATGCTGGCTCAGGTGCAAGTGGTGAATAT 2217
Qy      248 ArgPheIysGlyValIysTyrPheArgLeuAnthrLeuAlaSerIleuGlyTyrValVal 267
Db      2218 CGGTTTAAAGAGGATCAAGATATTCGCTGAAATACCTAGCTCTCTAGGTATGAGGTT 2277
Qy      268 ValValIleAspAspArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
Db      2278 GTAGGAAATAGACACAGGGGATCTGTGACCGAGGGCTTTAAATTTGAAGCGCCCTTTAA 2337
Qy      288 TyrLysMetGlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnTyrPheAlaSer 307
Db      2338 TATATAATGGGTCAATATGAAATTTGACGATTCAGGTGGAAGACTCCAAATATCTAGCTTCT 2397
Qy      308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyrGlyTyr 327
Db      2398 CGATATGATTTTCATGACTTAGATCGTGTGGGCATCCAGCGCTGGTCTTATGAGAGATTC 2457
Qy      328 LeuSerIleuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db      2458 CTCTCCCTGATGGCATTTATGACAGAGTCAAGATATCTTCAGGGTGTGCTATTCTGGGGCC 2517
Qy      348 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
Db      2518 CCAATCACTCTGTGGATCTTCTATGATACAGATACAG 2556

RESULT 6
US-09-976-674-12
; Sequence 12, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Jumeau, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPLIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 12
; LENGTH: 4829
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-12

Alignment Scores:
Pred. No.: 9,6e-223 Length: 4829
Score: 1816,50 Matches: 360
Percent Similarity: 67,29% Conservative: 0
Best Local Similarity: 67,29% Mismatches: 0
Query Match: 93,97% Indels: 175
Dbs: 10 Gaps: 1

US-10-070-464-7 (1-360) x US-09-976-674-12 (1-4829)
Qy      1 GluGluAspAlaArgSerIleGlyValAlaThrPheValIleuGlnGluGluPheAspArg 20
Db      943 GAAAGAGATGCCAGATACAGCTGAGTGGCTACTTTGTTCTCCAGAGAAATTTGATAGA 1002
Qy      21 TyrSerGlyTyrTyrTrpCysProIysAlaGluThrThrProSerGlyGlyLysIleu 40
Db      1003 TATTCGTGCTATTTGTGTGGTCCAAAGAGCTGAACACACTCCAGTGTGTAAATTTCTT 1062
Qy      41 ArgIleuTyrGluGluAspGluSerGlyValGlnIleIleHisAlaThrSerPro 60
Db      1063 AGAATTCATATGAGAAATATGATGATCTGAGGTGAATATTCATGATCCCT 1122
Qy      61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
Db      1123 ATGTTGAAACAGAGAGGAGATTCCTTCCTTAAACAGGTACAGCAATCT 1182
Qy      81 LysValThrPheLysMetSerGlnIleMetIleAspAlaGlyArgIleIle----- 98
Db      1183 AAAGTCACTTTAAAGATGTCAGAAATATATGATTTGATGTCGAAGAGATCATATGATGTC 1242
Qy      98 ----- 98
Db      1243 ATATGTAAGAACTAATTCACCTTTTGATTTCAATTTGAAGAGTTGAATATATATGTC 1302
Qy      98 ----- 98
Db      1303 AGAGCTGATGACCTCTGAGGAAATATGCTTGGTCCATCTACTAATGCTCCAG 1362
Qy      98 ----- 98
Db      1363 ACTCGCTACAGATATGTTGATCTCACCTGAATTTATTCACAGTAGAAGATATGTT 1422
Qy      98 ----- 98
Db      1423 ATGAAAGCAGAGACTCATTTGATGATGCTGATTCGTGACGCCACTAATATATCTAT 1482
Qy      98 ----- 98
Db      1483 GAAAGAAACAGACATCTGATTAATATATCATGACATCTTCATGTTTTCCTCCCAAGT 1542
Qy      98 ----- 98
Db      1543 CACGAAAGAGAAATTGAGTTTATTTTGGCTCTGAATGCAAAAACAGTTTCGTCAATTTA 1602
Qy      98 ----- 98
Db      1603 TACAAATATCATCTATTTTAAAGAAACAAATATTAACGATCCAGTGTGGCTGCT 1662
Qy      98 ----- 98
Db      1663 GCTCCAGATTTCAAGTCTCTTCAAAAGAGAGATGCAATTAACAGTGTGAATGG 1722
Qy      99 -----ValAspGluValArgArgLeuValTyr 1707
Db      1723 GAAGTTCTTGGCCGGCATGATCTAATATCCAAAGTTGATGAAGTCAAGAGGCTGATAT 1782
Qy      108 PheGluGlyThrLysAspSerProLeuGlnHisIleuTyrValValSerTyrValAsn 127
Db      1783 TTTGAAGGACCAAAAGACTCCCTTTAGGACATCACTGTACTGATCTGATTCATGCTAAAT 1842
Qy      128 ProGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
Db      1843 CTTGAGAGGTACAAAGGCTGTGACCTGACCTGCTACTCATTTCTTGCTGATCACTGAC 1902
Qy      148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db      1903 CACTGTGACTTCTTATATAGTATAGTATGTAACCAAGAAATCCACACTGTGTCTCTT 1962
Qy      168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrAlaThr 187
Db      1963 TACAAGCTATCAAGTCTGAAGATGACCCCACTTGCAAAACAAAGGAATTTTGGGCAAC 2022
Qy      188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 207
Db      2023 ATTTTGGATTCAGCAGGTCCTCTCTGACTATCTCTCTCCGAAATTTTCTTTTGA 2082
Qy      208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db      2083 AGTACTACTGATTTACATTTGATGAGATGCTCTCAAGCCCTCATGATCTACAGCTGGA 2142
Qy      228 LysLysTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGlnLeuValAsn 247
Db      2143 AAGAAATATCTTACAGTGTCTGTTCATATATGATGTGTCTCTCAGGGCAGTTGTGAATA 2202
Qy      247 snArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValV 267
Db      2203 ATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACGCTCTTACGTTATGTGG 2262
```

QY 267 aIvalValIleaspsnarqglyserCyshlearglyleuLysPhegluGlyAlaPheL 287
Db 2263 TTGTAGTATAGACAAACAGGGGATCCTGTCACCGAGGGCTTAATTGAGCGGCTTTA 2322
QY 287 ySTyLyMeGlvgInIleGluIleaspsnGlyValIgluGlyleuGlnTyLeuAlaAs 307
Db 2323 AATATAAATGGGTCAAAATAGAAATTAGACATCAGGTGGAGAGACTCCAAATATCTAGCTT 2382
QY 307 eTargTyAspPheIleaspleuasparqValIglYleHsglyTyPserTyrglyYT 327
Db 2383 CTGGATATGATTTGATTTGATCTTAGATCGTGGGATCCACGGCTGCTCTATGAGGAT 2442
QY 327 yTLeuSerleuMeLaleuMeCglnArgSerAspIlePheargValAlaIleAlaGlyA 347
Db 2443 ACCCTCCCTGATGGCATTAAAGCAGAGGTCAAGATATCTTCAGGGGTGCTATTGCTGGGG 2502
QY 347 laProlValThleuTrpIlePheTyAspTrpGlyTyTrpThr 360
Db 2503 CCCAGTCACTCTGTGATCTTCTATGATACAGGATACAG 2543

RESULT 7

US-09-976-674-22
; Sequence 22, Application us/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 4685
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-22

Alignment Scores:
Pred. No.: 8,42e-185 Length: 4685
Score: 1522.00 Matches: 312
Percent Similarity: 58.54% Conservative: 0
Best Local Similarity: 58.54% Mismatches: 1
Query Match: 78.74% Indels: 221
DB: 10 Gaps: 2

US-10-070-464-7 (1-360) x US-09-976-674-22 (1-4685)

QY 1 GIuGIuAspAlaArgSerIaGlyValAlaThrPheValIleuGlnGluIuPheAspArg 20
Db 943 GAAGAAGATGCCAGATCACTGGAAGTCCTACCTTTGTTCTCCAGAAATAATTGATGA 1002
QY 21 TySergLyTyTrpTyPserProLyAlaGluThrThrProSergLyGlyLySleu 40
Db 1003 TATTCGCTATTGGGTGTCGCAAAAGCTGAACAACATCCAGGTGGTGAATAATTTCTT 1062
QY 41 ArgIleLeuTyGlnGluAsnAspGluSerGluValGluIleIleHisValThrSerpPro 60
Db 1063 AAAAAATCTATATGAAGAAATATGATCTGAGGTGGAATATTCATATCATCCCT 1122
QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyTrpLySThrGlyThrAlaAsnPro 80
Db 1123 ATGTTGGAACAACAGAGGGCAATTCATCCGTTATCCTAATAACGGTACACCAATCCT 1182
QY 81 LysValThrPheLysMetSergLulMetIleAspAlaGluGlyArgIleIle----- 98
Db 1183 AAAGTCACTTTTAAATGTCAGAAATTAATGATTGATGTCGAAGAGATCATAGATGTC 1242

QY 98 ----- 98
Db 1243 ATAGATAAGAACTAATTCAACTTTGAGATTCATTATTGAAGAGTTGAATATATTGCC 1302
QY 98 ----- 98
Db 1303 AGAGCTGATGACCTCTGAGGAAATATGCTTGCTGCATCTACTAGATGCTCCAG 1362
QY 98 ----- 98
Db 1363 ACTGCTACAGATAGTGTGATCTCACCTGAATATTATTATCCAGTAGAAGATGATGTT 1422
QY 98 ----- 98
Db 1423 ATGAAAGCAGAGACTGATTGAGTCAGTCCGATCTGTGAGCCCACTAATTATCTAT 1482
QY 98 ----- 98
Db 1483 GAAGAAACAGACATCTGATTAATATCCATGACATCTTTGATGTTTCCCAAGT 1542
QY 98 ----- 98
Db 1543 CACGAAGAAATAGATTATTTTGGCTGTAATGCAAAACAGTTCCGTCATTTA 1602
QY 98 ----- 98
Db 1608 TACAAATTAATCTATTATTTAAAGAAACAAATTAACGATCCAGTGGTGGCTGCT 1662
QY 98 ----- 98
Db 1663 GCTCAAGTATTCAGATGTCCTATCAAAAGAGATAGCAATTAACAGTGTGATAG 1722
QY 99 -----ValaspGluValArgLeuValTy 107
Db 1723 GAAGTCTTGGCCGCATGATCTAATATCCAAAGTTGATGAAGTCAAGAGCTGGATAT 1782
QY 108 PhegluGlyThrIleAspSerProleuGlnHshSleuTyValValSerTyValaAsn 127
Db 1783 TTTGAAGGACCAAGAGACTCCCTTTAAGACATCCGTAGAGTACGTACGTACGTAAAT 1842
QY 128 ProGlyIuValThrArgLeuThrAspArgGlyTySerHisSerCysCysIleSergIn 147
Db 1843 CCTGAGAGGTACCAAGCTGACCTGACCGTGGCTACTGACATCTTCTGCGATCAGTCAG 1902
QY 148 HisCysAspPhePheIleSerLyTySerAsnGlnLyAsnProHisCysValSerleu 167
Db 1903 CACTGTGACTCTTTATATAGTAAGTATAGTAACTCAGAAAGATCCACACTGTGTGCTCTT 1962
QY 168 TyLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
Db 1963 TACAAAGCTATCAAGTCTGAAAGATGACCACTTCGCAAAACAAAGAAATTTGGGCCACC 2022
QY 2023 IleLeuAspSerAlaGlyProleuProAspTyTrpThrProGluIlePheSerPheGlu 207
Db 2023 ATTTGATTCAGT----- 2036
QY 208 SerThrThrGlyPheThrLeuTyrglyMetLeuTyTrpHisAspLeuGlnProGly 227
Db 2036 ----- 2036
QY 228 LysLyTyProThrValIleuPheIleTyrglyGlyProGlnValIgluLeuValAsnAsn 247
Db 2037 -----CTCAAGTCAAGTGGTGAATAAT 2060
QY 248 ArgPheLyGlyValIleTyPheArgLeuAsnThrIleuAlaSerLeuGlyTyValVal 267
Db 2061 CGGTTTAAGGAGCAAGATATCCGCTTGAATACCTAGCCTCTCTAGGTTATGAGTT 2120
QY 268 ValValIleaspsnarqglyserCyshlsargglyleuLysPhegluGlyAlaPheLys 287
Db 2121 GTAGTATAGACAAACAGGGGATCTGTACCGAGGGCTTAATTTGAAGGGCGCTTTAAA 2180

```

QY      288  TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
DB      2181  TATAAATGGGTCAAAATAGAAATTGACGATCAGGTGGAAAGACTCCAAATTTCTAGCTTCT 2240
QY      308  ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPseTyrGlyIleValTyr 327
DB      2241  CGATTAAGATTCTTATGCTAGATCGATGCGGCAATCCAGCGCTGCTCTTATGAGAGATAC 2300
QY      328  LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
DB      2301  CTCTCCCTGATGGCATTTAATGACAGAGTCAATCTTCAGGGTTGCTATGCTGGGGCC 2360
QY      348  ProValThrLeuThrIlePheTyrAspThrGlyTyrThr 360
DB      2361  CCAGTCACTCTGTGGATCTTCTATGATACGAGATACAG 2399

RESULT 8
US-09-976-674-20
; Sequence 20, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Akisanya, Karen
; APPLICANT: Qi, Steve
; APPLICANT: Riviere, Pierre
; APPLICANT: Jannier, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 4676
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-20

Alignment Scores:
Pred. No.:      9 74e-185      Length:      4676
Score:          1521.50      Matches:      309
Percent Similarity: 57.97%      Conservative: 0
Best Local Similarity: 57.97%      Mismatches: 1
Query Match:      78.71%      Indels:      224
DB:              10      Gaps:      2

US-10-070-464-7 (1-360) x US-09-976-674-20 (1-4676)
QY      1  GluGluAspAlaArgSerAlaGlyValAlaThrPheValIleuGlnGluIleuPheAspArg 20
DB      943  GAAGAAGATGCCAGATCAGCTGAGTCCGTAACCTTGTCTCCAGAAGAATTGATAGA 1002
QY      21  TyrSerGlyTyrThrProCysProLysAlaGluThrProSerGlyGlyLysIleLeu 40
DB      1003  TATTTCGGCTATTGGGTGTCCTCAAAAGCTGAAACCACTCCAGTGTGGTAAATTTCTT 1062
QY      41  ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValIleThrSerPro 60
DB      1063  AGAATTCTATATGAAGAAATGATGATCATCGAGTGGAAATTTATCTATCATCCCT 1122
QY      61  MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
DB      1123  ATGTGTGAACAAGAGGAGGAGATTCATTCCTTATCTTAAACAGGTACGCAAAATCTCT 1182
QY      81  LysValThrPheLysMetSerGluIleMetIleAspAlaGluGluValArgIleIle----- 98
DB      1183  AAAGTCACTTTTAAAGATGTCAGAAATATATGATGATGCTGAGAGAGATCATAGATGTC 1242
QY      98  ----- 98
DB      1243  ATAGATAAGGAACATAATTCACCTTTGAGATTCATTTGAAGAGATTTGATATATTGCC 1302

```

```

QY      98  ----- 98
DB      1303  AGAGCTGATGAGACTCTCGAGGAAATATGCTTGTGTCATCTACTAGATCGCTCCAG 1362
QY      98  ----- 98
DB      1363  ACTCGCTACAGATAGTGTGATCTGACCTGAATTTATATCCAGTAGAAGATGATGTT 1422
QY      98  ----- 98
DB      1423  ATGAAAGGAGAGACTCATGTGAGTCCCTGATTCGTGAGCCACTAATTAATCTAT 1482
QY      98  ----- 98
DB      1483  GAAGAAACAACAGACATCGATATAATATCATGATCTTTCATGTTTTCCTCCAAAGT 1542
QY      98  ----- 98
DB      1543  CACGAAGAGAAATGAGTTTATTTTGGCTCTGAATGCAAAACAGGTTCCGTCAATTA 1602
QY      98  ----- 98
DB      1603  TACAAAATTACATCTATTTTAAAGAAAGCAATATTAACGATCCAGTGGGCTGCCT 1662
QY      98  ----- 98
DB      1663  GCTCCAAAGTATTTCAAGTGTCTCTATCAAGAGAGATAGCAATTACAGTGTGAATGG 1722
QY      99  ----- ValAspGluValArgArgLeuValTyr 107
DB      1723  GAAGTCTTGGCCGGCATGATGATTAATATTCGAATGATAGCAAGAGCTGGATATAT 1782
QY      108  PheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValIleSerTyrValAsn 127
DB      1783  TTGAAGGACCAAAAGACTCCCTTTAGAGCATCACTGTACGTAGCATTTAGTAAAT 1842
QY      128  ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlySerGln 147
DB      1843  CTGGAAGAGGTGACAAAGCTGACCTGACCGGTACTACATCTTCTGTCATCAGTCAG 1902
QY      148  HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
DB      1903  CACTGAGACTCTTATTAAGTATAGTATAGTAAACAGAGATCCACCTGTGTCTCTT 1962
QY      168  TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheIleThr 187
DB      1963  TACAGCTATCAAGTCTCTGAAGATGACCCACTGTCAGAAACAAAGAAATTTGGCCACC 2022
QY      188  IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 207
DB      2023  ATTTGGATTCACAGAGTCTCTCTCTGACTATATCTCCAGAAATTTCTTTTGAA 2082
QY      208  SerThrThrGlyPheThrLeuTyrArgIleMetLeuTyrLysProHisAspLeuGlnProGly 227
DB      2083  AGTACTACGTGATTTACATTTGATGAGATCTCTTACAGGCTCATGATCTTACAGCTTGA 2142
QY      228  LysLeuTyrPheProThrValIleuPheIleTyrGlyGlyProGluValGlnLeuValAsn 247
DB      2143  ABAATAATCTTACTCTGTGCTTCATATAGTGTG----- 2180
QY      248  ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
DB      2180  ----- 2180
QY      268  ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
DB      2180  ----- 2180
QY      288  TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
DB      2181  -----GGTCAAAATAGAAATTTGACGATCAGGTGGAAAGACTCCAAATTTCTAGCTTCT 2221
QY      308  ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPseTyrGlyIleValTyr 327

```



```

Db      2232 CGATTGATTTCATTAGTACGATCGTGGGATCCAGCGCTGCTTATGAGATAC 2291
Qy      328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db      2292 CTCCTCCCTGATGACATTAAAGCAGAGGTGAGATATCTTCAGGGGTCTGATGCTGGGGCC 2351
Qy      348 ProValThrLeuThrIlePheThrAspThrGlyTyrThr 360
Db      2352 CCAGTCACCTCTGTGGATCTTCTATGATACAGGATACAG 2390

RESULT 9
US-09-976-674-8
; Sequence 8, Application US/0976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 4523
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-8

Alignment Scores:
Pred. No.: 4,69e-147 Length: 4523
Score: 1229.00 Matches: 258
Percent Similarity: 48.41% Conservative: 0
Best Local Similarity: 48.41% Mismatches: 1
Query Match: 63.58% Indels: 275
DB: 10 Gaps: 2

US-10-070-464-7 (1-360) x US-09-976-674-8 (1-4523)
Qy      1 GLUAluAspAlaArgSerIleGlyValAlaThrPheValLeuGlnGluIlePheAspArg 20
Db      943 GAAGAGATGCCAGATCAGCTGAGTGCCTACCTTCTCTCCAGAGAAATTTGATAGA 1002
Qy      21 TyrSerGlyTyrTyrProGlySerProGlyValGlnThrProSerGlyGlyIleLeu 40
Db      1003 TATTTCTGGCTATGTGGTGTCCAAAGCTGAAACAACTCCAGTGTGTAAATTTCTT 1062
Qy      41 ArgIleLeuTyrGluGlnuAspGluSerGluValGlnIleIleHisValThrSerPro 60
Db      1063 AGAATTTATATGAAGAAATGATGATCTGAGGTGAAATATATCTCATGTTACATCCCT 1122
Qy      61 MetLeuGlnThrArgArgAlaAspSerPheArgTyrProGlySerThrGlyThrAlaAspPro 80
Db      1123 ATGTTGAAAACAAGAGGAGGAGATTCATTCGGTATCTTAAACAGGTACAGCAATCCT 1182
Qy      81 LysValThrPheIleMetSerGluIleMetIleAspAlaGluGlyArgIleIle 98
Db      1183 AAAGTCACTTTTAAAGATGACGAATATATGATGATCTGAAGGAAGATCATATGATGTC 1242
Qy      98 2178
Db      1243 ATGATRAAGAACTAATTCACCTTTGAGATTCTATTGAAGAGCTTGAATATATTGCC 1302
Qy      98 2178
Db      1303 AGAGCTGATGACTCCTGAGGAAATATGCTGTGTCATCTCACTAGATCGCTCCAG 1362
Qy      98 2178

```

```

Db      1363 ACTCGCTACAGATAGTGTGATCTCACCTGATATTATTATCCAGTAGAAGATGTT 1422
Qy      98 2178
Db      1423 ATGAAAGCAGAGACTCATGATGACGTGCTGATTTCTGTGACGCCACTAATTATCTAT 1482
Qy      98 2178
Db      1483 GAAAGAAACAAGACATCTGATTAATATTCATGACATCTTTCATGTTTTCGCCCAAGT 1542
Qy      98 2178
Db      1543 CACGAAGAGAAATGATTTATTTTGGCTGTGAATGCAAAACAGGTTTCGTCATTTA 1602
Qy      98 2178
Db      1608 TACAAATTAATCATCTATTTTAAAGAAAGCAATATTAACATCCAGTGTGGCTGCC 1662
Qy      98 2178
Db      1663 GCTCCAGATGATTTCAAGTGTCTCTATCAAAAGAGAGATAGCAATTACAGTGTGAATGG 1722
Qy      99 2178
Db      1723 GAAGTCTTGGCCGCGATGATCTAATATTCAAAGTTGATGAAGTCAGAAAGCTGTATAT 1782
Qy      108 PheGluGlyThrIleAspSerProLeuGlnHisIleLeuTyrValAlaSerTyrValAsn 127
Db      1783 TTTGAAGGACCAAAACCTCCCTTTAGAGCAACCTGTACCTGATGCAATTCGTAAT 1842
Qy      128 ProGluGlyValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
Db      1843 CCTGAGAGGTGACAAAGCTGACTGACCGGTGACTACATCTTCTGCTGATCAGTCAG 1902
Qy      148 HisCysAspPhePheIleSerIleTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db      1903 CACTGAGACTTCTTTATAGTATGATGATGATGACCAAGAAATCCACACTGTGTGCTT 1962
Qy      168 TyrIleLeuSerSerProGluuAspAspProThrCysIleThrIleGluPheThrPalaThr 187
Db      1963 TACAACTATCAAGTCTTGAAAGTGAACCAACTGCAAAACAAAGAAATTTGGGCCACC 2022
Qy      188 IleLeuAspSerIleGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db      2023 ATTTGATTCAGCAGAGTCTCTCTGATATATCTCCAGAAATTTCTCTTTGAA 2082
Qy      208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleProHisAspLeuGlnProGly 227
Db      2083 AGTACTACGTGATTTACATTTGATGGGATGCTCTACAGCCTCATATCTACAGCCCGGA 2142
Qy      228 LysIleTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
Db      2143 AAGAAATATCTTCTGTGCTGCTTCATATATATGGGT 2178
Qy      248 ArgPheLysGlyValIleTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db      2178 2178
Qy      268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
Db      2178 2178
Qy      288 TyrIleMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db      2178 2178
Qy      308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyGlyTyr 327
Db      2178 2178
Qy      328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347

```

Db 2179 -----CGGTT-GCTATTGCTGGGCC 2198
Qy 348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
Db 2199 CCAAGTCACTCTGTGGATCTTCTATGATACAGGATACAG 2237

RESULT 10

US-09-976-674-4
; Sequence 4, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2617
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-4

Alignment Scores:

Pred. No.: 2,466-146 Length: 2617
Score: 1220.50 Matches: 236
Percent Similarity: 54.31% Conservative: 54
Best Local Similarity: 44.19% Mismatches: 69
Query Match: 63.14% Indels: 175
Gaps: 2

US-10-070-464-7 (1-360) x US-09-976-674-4 (1-2617)

Qy 2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGluPheAspArgTyr 21
Db 710 GATGACCCCAAGTGTGGGGGTGGCCACCTTCGTCAACAGGAAGATTGCACCGCTTC 769
Qy 22 SerGlyTyrTrpCysProLysAlaGluThrThrProSerGlyGly---LysIleLeu 40
Db 770 ACTGGTACTGTGTGGTCCCAACAGCTCTCGGAAGGTTCAGAGGGCTCAAGACGCTG 829
Qy 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
Db 830 CGAATCCCTGTAAGAGAGAGTGAATGATCCGAGGTGAGAGTCAATCAGTCCCTCTCT 889
Qy 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
Db 890 GCGCTAGAAAGAAAGAAAGACGAGACTCGATCGGTACCCAGAGACAGCAAGATCCC 949
Qy 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
Db 950 AAGATTGCTTGAAGCTGAGTTCAGACTGACACGCGGCAAGATGCTTCGACC 1009
Qy 98 ----- 98
Db 1010 CAGAGAAAGAGCTGTGTGACGCTTCAGCTCCGCTTTCCGGAAGGTGAGATCACTGCC 1069
Qy 98 ----- 98
Db 1070 AGGCGCGGTGAGACCCGGGATGCAAAATAGCCTGGGCGCATGTTCTGSAACGCGCCAG 1129
Qy 98 ----- 98
Db 1130 CAGTGGCTCAGCTGCTCTCTCCCGCGCGCTGTTCAATCCGAGAGACAGAGATGAG 1189
Qy 98 ----- 98

Db 1190 GAGACGCGGTAGCTCTGCGCAGAGCTGTCCCAAGAAATGTCACGCCGATGTGGTAC 1249
Qy 98 ----- 98
Db 1250 GAGAGGTCAACCAAGTGTGATCAATGTCATGATCATTTCTATCCCTTCCCAATCA 1309
Qy 98 ----- 98
Db 1310 GAGGAGAGACAGACTCTGCTTTCTCCGCGCATGATGATGAGACAGCGCTTTCGCAT 1369
Qy 98 ----- 98
Db 1370 TTGTAACAAGTCACCGCGTTTAAATCCAGGGCTACGATGAGAGCCCTTACAC 1429
Qy 98 ----- 98
Db 1430 CCGGGGAAGATGAATTAAGTCCCACTTAAGAGAGATGCTCTGACCGCGGTGA 1489
Qy 99 -----ValAspGluValArgLeuVal 106
Db 1490 TGGAGGTTTGGCGAGCAGCGCTCAAGATCTGGGTCAATGAGAGACCAAGCTGTGTG 1549
Qy 107 TyrPheGluGlyThrLysAspSerProLeuGluHisIleuTyrValValSerTyrVal 126
Db 1550 TACTTCAGGGCACAAGACACCGCGCTGAGACACACCTCTACGTGCTGCTGATGAG 1609
Qy 127 AsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSer 146
Db 1610 GCGCGCGCGAGATGCTAGCTGCTGACACAGCCCGCTCTCCATAGCTGCTCCATGAGC 1669
Qy 147 GluHisCysAspPhePheIleSerLysTyrSerAsnGluLysAsnProHisCysValSer 166
Db 1670 CAGAACTTGCACATGTTGTCAGGCACATACAGACGTCGAGACAGCGCGCTGCGTGCAC 1729
Qy 167 LeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrAla 186
Db 1730 GCTACAAAGCTGAGGGGCCCGACGACGACCTCTGCAACAGACAGCCCGCTTCTGGGT 1789
Qy 187 ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPhe 206
Db 1790 AGCATGATGAGAGCGACGACGCTGCGCCCGGATTAATGTTCTCCAGAGATCTTCCATTTC 1849
Qy 207 GluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnPro 226
Db 1850 CACAGCGCTCGAATGTCGCGCTCTACGCGATGATCTACAGCCCAAGCGCTTGACGCA 1909
Qy 227 GlyLysLysTyrProThrValLeuPheIleTyrGlyLysProGluValGluLeuValAsn 246
Db 1910 GGGAAAGAGCACCCACGCTCTTGTATATGAGAGCCCCCAAGTGCAGCTGTGTGAT 1969
Qy 247 AsnArgPheLysGlyValIleTyrThrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrVal 266
Db 1970 AATCTCTTCAAGGATCAATATCTGCGGCTCAACACACTGCGCTTCCCTGGGCTACGCC 2029
Qy 267 ValValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluValAlaPhe 286
Db 2030 GTGGTTGATTGACGCGAGGGGCTCTCTGACGAGAGGGCTTCGGTTGGAAGGGCCCTG 2089
Qy 287 LysTyrLysMetGlyGluIleGluIleAspAspGluValGluLysLeuGlyThrLeuAla 306
Db 2090 AAAAACCAATGGGCGAGTGAATTCAGGACCAAGGTGAGGGGCTTCAGTTCTGTGGC 2149
Qy 307 SerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGly 326
Db 2150 GAGAAATATGCTTCACTGACACTAGCGGATTCGATCCATTCATGCTGCTCTTACGGGGG 2209
Qy 327 TyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 346
Db 2210 TTCTCTCTGCTCATGCGGCTAATCAACAAGCCCAAGGTTCAAGTGGCATGCGGGGT 2269
Qy 347 AlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
Db 2270 GCCCGGTACCGTGTGATGGCTTACGACAGAGGTACCT 2311

RESULT 11

US-09-976-674-28
; Sequence 28, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Juhien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPEIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 4219
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-28

Alignment Scores:

Pred. No.:	5,24e-146	Length:	4219
Score:	1220.50	Matches:	236
Percent Similarity:	54.31%	Conservative:	54
Best Local Similarity:	44.19%	Mismatches:	69
Query Match:	63.14%	Indels:	175
DB:	10	Gaps:	2

US-10-070-464-7 (1-360) x US-09-976-674-28 (1-4219)

```
QY      2  GluSpAlaArGSeRrAlaGlyValAlaThrPheValLeuGInGluPheAspArgTyr 21
DB      1066 GATGACCCCAAGTCTGGGGGTGGGCCACTTCGTCAACAGAAAGTTGACCGCTTC 1125

QY      22 SerGlyTyrTPProGlyProGlyAlaGluThrProSerGlyGly--LysIleLeu 40
DB      1126 ACTGGTACTGGTGGTGGCCCAAGCCCTCTGGAAAGTTGAGAGGCTTCAAGACGCTG 1185

QY      41 ArgIleuTyrGInGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
DB      1186 CGATCTGTATGAGAAAGTGAAGTGAAGTCCGAGGTGAGGTCATTCAGTCCCTCTCT 1245

QY      61 MetLeuGluThrArGArGAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
DB      1246 GCGCTAGAAAGAAAGAAAGACGACTCGTATCGTAAACCCAGAGACGAGCAAGAAATCCC 1305

QY      81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyValGlyIle----- 98
DB      1306 AAGATGCTTGAAGCTGGCTGAGTTCAGACTACAGCCAGGGCAAGATGCTCGACC 1365

QY      98 ----- 98
DB      1366 CAGGAGAAGAGCTGTGACAGCCCTTCACTCGCTGTCCCGAAGTGAATACATGCC 1425

QY      98 ----- 98
DB      1426 AGGGCCGGGTGGACCCGGGATGCAAAATACGCTGGGCCATGTTCTTGACCCGCCAG 1485

QY      98 ----- 98
DB      1486 CAGTGGCTCCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1545

QY      98 ----- 98
DB      1546 GAGCAGCGGCTAGCGCTTGCCAGAGCTGTCCCGAGAAATGTCAGCCGTATGTGTGAC 1605

QY      98 ----- 98
DB      1606 GAGAGGTACCAACGTCTGGATCAATGTTCATGACATCTTATCTCTCTCTCTCTCTCTCT 1665
```

QY 98 ----- 98

DB 1666 GAGGAGAAGACGAGCTCTGCTTCTCCGCCAATGAAATGCAAGACCGCTTGTCCAT 1725

QY 98 ----- 98

DB 1726 TTGTACAAAGTCACCGCGCTTTTAAATCCAGGGCTTACGATGAGAGCCCTTCAGC 1785

QY 98 ----- 98

DB 1786 CCCGGGAAGATGAATTTAAGTGCCCATTAAGGAAGAGATTGCTGTGACCGGTGA 1845

QY 99 -----ValaspGluValArgIleVal 106

DB 1846 TGGAGGTTTGGCGAGCAGGCTCCAGATCTGGGTCATGAGAGCAAGCTGGTG 1905

QY 107 TyrPheGluGlyThrLysAspSerProLeuGluHisIleuTyrValValSerTyrVal 126

DB 1906 TACTTCAGGGCACCAGACACCGCGTGAGCACCACTTACGTGTGAGTCACTATGAG 1965

QY 127 AsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSer 146

DB 1966 GCGCGCGGAGATGTACGCTTACCACCGCGGCTTCTCCATGCTGTCCATGAGC 2025

QY 147 GluHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSer 166

DB 2026 CAGAACTTCGACATGTTGTGTAAGCACTACAGCGGTGAGCAGCCGCTGTGTCAC 2085

QY 167 LeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAla 186

DB 2086 GTTCAAGAGCTGACGGCCCGGACGACGCCCTTCGACAAAGCGCCCGCTTCTGGCT 2145

QY 187 ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPhe 206

DB 2146 AGCATGATGAGGAGCAGCAGCTGCCCCGAGATTAATGTTCTCCAGAGATCTTCATTC 2205

QY 207 GluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnPro 226

DB 2206 CACAGCGCTCGAGTGGCGCTCTACGCAATGATCAAGCCCAAGCCTTCAGCA 2265

QY 227 GlyLysLysTyrProThrValIleuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 246

DB 2266 GGGAAAGACACCCACCGCTCTTTGTATATGAGGCCCCCGAGTGCACCTGTGAAAT 2325

QY 249 AsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrVal 266

DB 2326 AACTCTTCAAAAGCATCAAGTACTTGGCTCAACACACTGGCTCTCTGGGCTAGGCC 2385

QY 267 ValValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGlyValAlaPhe 286

DB 2386 GTGGTTGTGATGACGCGAGGGGCTCTGTACAGGAGGGCTTCGATTGCAAGGGGCTGTG 2445

QY 287 LysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAla 306

DB 2446 AAAAACCAATGGCCACGCTGAGATCGAAGACCCAGGTGAGGCTGTGCACTGTGACC 2505

QY 307 SerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheTyrGlyGly 326

DB 2506 GAGAAATGATGGCTTCACTGACCTGAGCGGATTCATCCATGAGCTGTGCTTCAAGGGGGC 2565

QY 327 TyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 346

DB 2566 TTCTCTCGCTCATGGGGCTTATCATCAAGCCCAAGGTGTTCAGAGTGGCATGCGGGGT 2625

QY 347 AlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360

DB 2626 GCCCGGTCAACCGTCTGATGAGCTTACGACACAGGGTACACT 2667

RESULT 12

US-09-976-674-24
; Sequence 24, Application US/09976674
; Patent No. US20020115843A1

```

: GENERAL INFORMATION:
: APPLICANT: Qi, Steve
: APPLICANT: Akinsanya, Karen
: APPLICANT: Riviere, Pierre
: APPLICANT: Junier, Jean-Louis
: TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPLV
: FILE REFERENCE: 70869
: CURRENT APPLICATION NUMBER: US/09/976,674
: CURRENT FILING DATE: 2001-10-12
: PRIOR APPLICATION NUMBER: US 60/240,117
: PRIOR FILING DATE: 2000-10-12
: NUMBER OF SEQ ID NOS: 61
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 24
: LENGTH: 4302
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-976-674-24

```

;; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
;; FILE REFERENCE: 70669
;; CURRENT APPLICATION NUMBER: US/09/976,674
;; CURRENT FILING DATE: 2001-10-12
;; PRIOR APPLICATION NUMBER: US 60/240,117
;; PRIOR FILING DATE: 2000-10-12
;; NUMBER OF SEQ ID NOS: 61
;; SOFTWARE: Patentin version 3.1
;; SEQ ID NO 36
;; LENGTH: 4180
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-976-674-36

Alignment Scores:
Pred. No.: 1,69e-143 Length: 4180
Score: 1201.00 Matches: 234
Percent Similarity: 55.09% Conservative: 53
Best Local Similarity: 44.91% Mismatches: 72
Query Match: 62.13% Indels: 162
DB: 10 Gaps: 3

US-10-070-464-7 (1-360) x US-09-976-674-36 (1-4180)

QY 2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluIlePheAspArgTyr 21
DB 1066 GATGACCCCAAGCTCGGGGTGGCCACCTTCTCATACAGAAAGTTCCAGCGCTTC 1125
QY 22 SerGlyTyrTrpTyrCysProIysAlaGluThrThrProSerGlyGly--LysIleLeu 40
DB 1126 ACTGGGTACTGGTGGTGGCCCAAGCCTCTCTGGAAAGTTCCAGAGGCGCTCAAGAGCGCTG 1185
QY 41 ArgIleLeuTyrGluGluIleAsnArgGluSerGluValGluIleIleHisValThrSerPro 60
DB 1186 CGATCTCTGATGAGAAAGTCGATGAGTCGAGTGGAGGTCAATTCAGTCCCTCTCTCT 1245
QY 61 MetLeuGluThrArgAlaAspSerPheArgTyrProIysThrGlyThrAlaAsnPro 80
DB 1246 GCGCTGAAAGAAAGAAAGAGACGACTCGTATCGGTACCCCAAGAGCGCAAGAAATCCC 1305
QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluIleArgIleIleValAsp 100
DB 1306 AAGATTGCTTGAACCTGGCTGAGTTCAGACTGACAGCAGGCGCAAGATCGTCTGAGCC 1365
QY 101 GluValArgArgLeu----- 105
DB 1366 CAGGAGAGAGAGCTGTGTGACAGCCCTTCACTGCTGTCCTCCGAAAGGTGAGTACATCGCC 1425
QY 105 ----- 105
DB 1426 AGGCGCGGTGACCCCGGATGGCAATAACGCTGGGCATGTTCTGTGACCGGCCCCAG 1485
QY 105 ----- 105
DB 1486 CAGTGGCTCCAGCTCGTCTCTCTCCCGGCGCTGTTCATCCCGAGCAGAGAATGAG 1545
QY 106 -----ValTyr 107
DB 1546 GAGCAGCGGTAGCTCTGCGCAGAGCTGTCCAGAAATGTCAGAGCCGATGTGTGTATC 1605
QY 108 PheGlu----- 109
DB 1606 GAGGAGGTCAACAAGCTGTGATCAATGTTGATGACATCTTATCCCTTCCCCCAATCA 1665
QY 109 ----- 109
DB 1666 GAGGAGAGAGCAGAGCTGTGCTTTTCCCGCAATGAATGCAAGACCGCTTGTGCAT 1725
QY 109 ----- 109
DB 1726 TTGTACAAAGTACCGCGCTTTAAATCCAGAGGCTAGCATTTGAGTGAAGCCCTTCAAG 1785
QY 109 ----- 109

DB 1786 CCGGAGAAAGATTAATTAGTCCCATTAAGAAAGACATTGCTGACACCGGTGAA 1845
QY 110 -----GlyThrLysAspSerProLeuGlnHis 119
DB 1846 TGGAGGTTTGGCGAGCAGCGCTCCAGAGGCAACCAAGACACCGCGCTGAGACAC 1905
QY 120 LeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgTyr 139
DB 1906 CTCTACGTGTGATGATGAGCGCGCGGAGATGTGCTTCAACCGCCCGGCTTC 1965
QY 140 SerHisSerCysCysIleSerGlnHisCysAspPheIleSerTyrSerAsnGln 159
DB 1966 TCCCATAGTGTCTCCATGAGCAGAACTTCGATGTCTGCGCATACACAGCAGCGTG 2025
QY 160 LysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCys 179
DB 2026 AGACAGCGCCCGCTGCTGACGCTCAAGCTCAAGCTGAGCGGCCCAAGACAGACCCCTGAC 2085
QY 180 LysThrLysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThr 199
DB 2086 AAGCAGCCCGCTTCTGGGCTAGCATGATGAGCAGCAGCAGCTCCCGCATTTATGTT 2145
QY 200 ProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyr 219
DB 2146 CTTCGAGATCTTCATTTCCACAGCGCTGAGATGTGCGGCTTCAAGCATGATCTAC 2205
QY 220 LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGly 239
DB 2206 AAGCCCAAGCCTTGACGACGAGAAAGAACCCACCGCTCTTTGATATGAGAGGC 2265
QY 240 ProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThr 259
DB 2266 CCCCAGTCCAGTGTGTGATTAATCTTCAAAAGCATCAAGTATGCGCTCAACACA 2325
QY 260 LeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCysHisArgGly 279
DB 2326 CTGGCTTCTCTGGGCTACCGCTGTGTGATTTGAGCGCAGGAGCTCTCTGACGAGGG 2385
QY 280 LeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAsnGlnVal 299
DB 2386 CTTCGTTGGAAGGGGCGCTGAAACCAATGGCGCAGGTGAGATCCAGACAGGTG 2445
QY 300 GluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 319
DB 2446 GAGGCGCTCAGTGTGTGCTGCGCAGAAATGAGCTTCATGACCTGAGCCGAGTTGCCATC 2505
QY 320 HisGlyTrpSerTyrGlyTyrLysSerLeuMetAlaLeuMetGlnArgSerAspIle 339
DB 2506 CATGCTGTCTCTACGGGCGCTTCTCTGCTCATGGGCTAATCCACAAGCCCCAGGTG 2565
QY 340 PheArgValAlaIleAlaIleAlaIleAlaProValThrLeuTrpIlePheTyrAspThrGlyTyr 359
DB 2566 TTCAAGGTGGCATCGCGGCTGCCCGGTGACCGCTGATGGCTAATCCAGCAGAGGTAC 2625
QY 360 Thr 360
DB 2626 ACT 2628

RESULT 14
US-09-976-674-34
Sequence 34, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akinsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976,674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,117

```

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 4263
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-34

Alignment Scores:
Pred. No.:      1,74e-143      Length:      4263
Score:          1201.00        Matches:      234
Percent Similarity: 55.09%      Conservative: 53
Best Local Similarity: 44.91%    Mismatches:  72
Query Match:     62.13%        Indels:      162
DB:              10           Gaps:         3

US-10-070-464-7 (1-360) x US-09-976-674-34 (1-4263)

QY      2  GluSepAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluPheAspArgTyr  21
      1066 GATGACCCCAAGTCTGCGGGTGTGGCCACCTTCGTCAATACAGAGAGAGTTGACCGCTTC  1125
QY      22  SerGlyTyrTrpTyrProGlyAlaGluThrThrProSerGlyGly---LysIleLeu  40
      1126 ACTGGGACTGGTGTGGCCCAACAGCTCTCGGAGGTTGAGAGGGCTCTCAAGAGCTG  1185
QY      41  ArgIleLeuTyrGlnGluAsnAspGluSerGluValGluIleLeuIleValThrSerPro  60
      1186 CGAATCCCTGATGAGAGAGTGAATGATGATCCAGATGGAGGTCAATTCAGTCCCTCTCT  1245
QY      61  MetLeuGlnThrArgAlaAspSerPheArgTyrProLysThrAlaAsnPro  80
      1246 GCGCTAGAGAAAGAGAGACGAGACTCGTATCGGTACCCAGAGACAGCAGCAAGATCCC  1305
QY      81  LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleLeuValasp  100
      1306 AAGATTGCTTGAAGCTGAGTCAAGTTCAGACTGACACGAGGAGAGATGCTTCGACC  1365
QY      101 GluValArgArgLeu-----  105
      1366 CAGAGAGAGAGAGCTGTGTGACAGCTTCAGCTCGCTGTTCCGAAAGTGAATACGCGC  1425
QY      105 -----  105
      1426 AGGGCCGGGTGAGACCCGGGATGACAAATACGCTGGGCATGTTCTGGACCGGCCAG  1485
QY      105 -----  105
      1486 CAGTGGCTCCAGCTGTCTCTCTCCCGCGCTGTTTCATCCGAGACAGAGAAATGAG  1545
QY      106 -----ValTyr  107
      1546 GAGCAGCGGCTAGCCTCTGCCAGAGAGCTGCCCGAGAAATGTCACGCCGTATGTGTAC  1605
QY      108 PheGlu-----  109
      1606 GAGGAGTTCACCAAGTGTGATCAATGTATGACATCTTATCCTTCCCAATCA  1665
QY      109 -----  109
      1666 GAGGAGAGAGAGAGCTGTCTTTTCCGCGCAATGATGACAGACCGGCTTCTGCAT  1725
QY      109 -----  109
      1726 TTGTACAAAGTCAACGCGGTTTAAATCCAGGGCTACGATTTGAGTGAAGAGCCCTTAC  1785
QY      109 -----  109
      1786 CCGGGGGAAGATTAATTAAAGTCCCATTAAGAGAGATTGCTTGACCAAGCGGTGA  1845
QY      110 -----GlyThrLysAspSerProLeuGluHisHis  119
      |||||

```

```

DB      1846 TGGAGGTTTGGCGAGGACAGGCTCCAGGGACCAAGAGACCGCCGCTGAGACCAAC  1905
QY      120  LeuTyrValIleSerTyrValAlaAsnProGlyGluValThrArgLeuThrAspArgLysTyr  139
      1906 CTCTACGTGTCAAGTATGAGCGCGCCGAGAGATGTAAGCTTCAACACCGCCGCTTC  1965
QY      140  SerHisSerCysValIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGln  159
      1966 TCCCATAGCTCTCCATAGGACAGAACTTCGACATGTCTGTCAGCACATACAGACGCGT  2025
QY      160  LysAsnProHisCysValSerLeuTyrTyrLysLeuSerSerProGluAspAspProThrCys  179
      2026 AGCAGCGGCGCTGTGTGACAGTCTACAGAGTGAAGCGGCCCGCAGACACACCCCTGCAC  2085
QY      180  LysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThr  199
      2086 AAGCAGCCCGCTTGTGGCTAGCATGATGAGAGAGCAGCAGTGGCCCCCGATTTATGTT  2145
QY      200  ProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyr  219
      2146 CTTCCAGAGATCTTCATTTCCACACGCGCTCGATGTGCGGCTCTACGCGCATGATCTAC  2205
QY      220  LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly  239
      2206 AAGCCCAAGCTTGTACAGCCAGAGAGAGACCCACCGTCTCTTGTATATGAGAGC  2265
QY      240  ProGlnValGlnLeuValAsnAspArgPheLysGlyValLysTyrPheArgLeuAsnThr  259
      2266 CCCCAGGTGACAGCTGTGAATTAATCTCTCAAGAGCATCAAGTACTTGGGGCTCAACACA  2325
QY      260  LeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGly  279
      2326 CTGGCTCTCTCGGGCTACGCGGTGTGATTAAGCGCAGAGGCTCTGTCAGCGAGAGG  2385
QY      280  LeuLysPheGlnGluGlyAlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGlnVal  299
      2386 CTTCGTTTGAAAGGGCCCTGAAAAACCAATGGCCAGGTGAGATGAGAGACCAAGGTG  2445
QY      300  GluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle  319
      2446 GAGGCGCTCGAGTGTGTGCGGAGAGATGAGCTTCAATGACCTGACCGGAGTTGGCATC  2505
QY      320  HisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIle  339
      2506 CATGGCTGTCTACGGGGGCTTCTCTCTGCTCAATGGGCTATCAACAGCCCAAGGTG  2565
QY      340  PheArgValAlaIleAlaGlyAlaProValThrLeuTyrPheTyrAspThrGlyTyr  359
      2566 TTCAAGTGGCCATGCGGGTGGCCCGGTCAACCGTGTGATGGCTTACGACACAGGGTAC  2625
QY      360  Thr 360
      2626 ACT 2628
DB

```

RESULT 15
 US-10-098-841-83
 ; Sequence 83, Application US/10098841
 ; Publication No. US20020197679A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Asundi, Vinod
 ; APPLICANT: Xu, Chongjun
 ; APPLICANT: Zhou, Ping
 ; APPLICANT: Ma, Yunding
 ; APPLICANT: Wang, Jian-Rui
 ; APPLICANT: Zhao, Qing A.
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Chen, Rui-hong
 ; APPLICANT: Wang, Dunrui
 ; APPLICANT: Wang, Zhiwei
 ; APPLICANT: Wehrman, Tom
 ; APPLICANT: Zhang, Jie

/ APPLICANT: Qian, Xiaohong B.
 / APPLICANT: Dmanac, Radoje T.
 / TITLE OF INVENTION: No. US20020197679A1 Nucleic Acids and
 / FILE OF INVENTION: Polypeptides
 / FILE REFERENCE: 784CIP2
 / CURRENT APPLICATION NUMBER: US/10/098,841
 / PRIOR FILING DATE: 2002-03-13
 / PRIOR APPLICATION NUMBER: 09/598,042
 / PRIOR FILING DATE: 2000-06-20
 / PRIOR APPLICATION NUMBER: 09/552,317
 / PRIOR FILING DATE: 2000-04-25
 / PRIOR APPLICATION NUMBER: 09/488,725
 / PRIOR FILING DATE: 2000-01-21
 / NUMBER OF SEQ ID NOS: 331
 / SOFTWARE: pc_fl_genes Version 1.0
 / SEQ ID NO: 83
 / LENGTH: 3262
 / TYPE: DNA
 / ORGANISM: Homo sapiens
 / FEATURE:
 / NAME/KEY: CDS
 / LOCATION: (121)..(2334)
 / US-10-098-841-83

Alignment Scores:

Pred. No.:	2,05e-141	Length:	3262
Score:	1183.50	Matches:	230
Percent Similarity:	55.34%	Conservative:	55
Best Local Similarity:	44.66%	Mismatches:	65
Query Match:	61.23%	Indels:	165
DB:	13	Gaps:	3

US-10-070-464-7 (1-360) x US-10-098-841-83 (1-3262)

QY 2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluPheAspArgTyr 21
 DB 499 GATGACCCCAAGCTGCGGAGTGTGGCCACCTTCGTATACAGGAAGATTGACCGCTTC 558
 QY 22 SerGlyTyrTrpProPheProLysAlaGluThrProSerGlyGly---LysIleLeu 40
 DB 559 ACTGGGTACTGGTGGGCCCCACAGCTCTCGGAAAGTTCCAGAGGAGCCCTCAAGACGCTG 618
 QY 41 ArgIleLeuTyrGlnGluAsnAspGluSerGluValGlnIleIleHisValThrSerPro 60
 DB 619 CGATCTGTATAGAGAAAGTCATGATCGAGGTGAGGTCACTTACGTCCTCTCTCT 678
 QY 61 MetLeuGluThrArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
 DB 679 GCGCTAGAAAGAAAGGAAAGCGACTGCTATCGGTACCCAGGACAGGACAGCAAGATCCC 738
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGlnGlyArgIleIle----- 98
 DB 739 AAGATTGCTTGAATCTGGCTGAGTTCAGACTGACAGCCAGGAGCAAGATGCTCTGACC 798
 QY 98 ----- 98
 DB 799 CAGAGAGAGAGCTGGTGCAGCCCTCAGCTGCTGTCCCGAAGGTGAGTACATCGCC 858
 QY 98 ----- 98
 DB 859 AAGGCGGGGTGACCCGGAGTGGCAATAACGCTGGGCAATGTTCTTGAGACGGCCCCAG 918
 QY 98 ----- 98
 DB 919 CAGTGGCTCAGCTGCT 978
 QY 98 ----- 98
 DB 979 GAGCAGCGGCTAGCTCTCTGACAGAGCTGTCCCGAGAAATGTCAGCGCTATGTGTATC 1038
 QY 98 ----- 98
 DB 1039 GAGGAGGTCAACCACTGTGATCAATGTTATGATGACATCTTATCTCTCTCTCTCTCTCT 1098

QY 98 ----- 98
 DB 1099 GAGGAGAGACAGACTGCTTCTTCTCCGCCAATGATGACAGACCGGCTTCTGCAAT 1158
 QY 98 ----- 98
 DB 1159 TTGTACAAAGTACCGCGCTTTAAATTCACAGGCTACGATTGAGTACGCCCTTCACG 1218
 QY 99 -----ValAspGluValArgArgLeu 105
 DB 1219 CCGGGGAAGTGAAGAGAGAGCTTACGATGCTATGTGGTCAATGAGAGACCAAGCTG 1278
 QY 106 ValTyrPheGlnGlyThrLysAspSerProLeuGlnHisIleLeuTyrValValSerTyr 125
 DB 1279 GTGTACTTCCAGGAGACCAAGACACCGCGCTGGAGACACCTCTACGTTGTCAGCTAT 1338
 QY 126 ValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysIle 145
 DB 1339 GAGCGCGCGCGAGATCGTACGCTCACACGCGCGCTTCTCCATAGCTGCTCCATG 1398
 QY 146 SerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysVal 165
 DB 1399 AGCCAGAACTTGCAGATGTTGTCAGCCACTACAGAGCGTGAAGACCGCGCGCTGCTG 1458
 QY 166 SerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyr 185
 DB 1459 CAGCTTACACCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTG 1518
 QY 186 AlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGlnIlePheSer 205
 DB 1519 GCTAGCATATGAGAGCAGC-----AAGATCTTCCAT 1551
 QY 206 PheGluSerThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGln 225
 DB 1552 TTCACACCGCGCTCGATGTGCGGCTCTACGCGCATATGATACAGCCCGCTTGCAG 1611
 QY 226 ProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuVal 245
 DB 1612 CCAGGAAAGAACACCCACCGCTCTTGTATATGAGGCGCCCGAGGTGACGCTGGTG 1671
 QY 246 AsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyr 265
 DB 1672 AATTACTCTTCAAAAGCATCAAGTACTTGGCGCTCAACACTGGCTCTCTCGGCTAC 1731
 QY 266 ValValValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGlnGlyAla 285
 DB 1732 GCGGTGTGTGATTGACCGGAGGCGCTCTGTACGCGAGGCGCTTCCAGAGGCGCC 1791
 QY 286 PheLysTyrLysMetGlyGlnIleGluIleAspAsnArgValGlnGlyLeuGlnIleLeu 305
 DB 1792 CTGAATAACCAATAGGCGCGAGGTGAGATCGAGACAGGTGAGGCGCTGCGATTCGTG 1851
 QY 306 AlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGly 325
 DB 1852 GCCGAGAGATATGGCTTATGACCTGAGCCGAGAGTTCATCATGCTGCTGCTGCTGCTG 1911
 QY 326 GlyTyrLysSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAla 345
 DB 1912 GGCTTCTCTCTCTATGGGCTAATCCACAGCCCGAGTTCATAGGTGCGCATCGCG 1971
 QY 346 GlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
 DB 1972 GGTGCGCGGTACCGCTGTGATGAGCTTACGACAGAGGTTACT 2016

Search completed: October 16, 2003, 09:10:05
 Job time : 309.152 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:24:12 ; Search time 2162.86 Seconds

(without alignments)
4045.398 Million cell updates/sec

Title: US-10-070-464-7

Sequence: 1 EEDARSAGVATFLQEEFDR.....RVAAIGAPVTMTFYDTGYT 360

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=x1h
-O=/cgn2.1/USPTO.spool/US10070464/runat.15102003.113553.24829/app.query.fasta.1.2652
-DB=EST -OPMT=fastp -SUFFIX=est -MINMATCH=0.1 -IOCPCT=0 -IOCPXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=Pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=Plco -NOM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464@cgn.1.1.8056@runat.15102003.113553.24829 -NCPU=6 -ICPU=3
-NO MMAP -LARGEBUFFER -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -NARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -Delop=6 -Delext=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_hiv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mus:*
23: em_gss_mam:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1.*

0 29 gb_gss2.*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB	ID	Description
1	1756.5	90.9	3143	11	AK016546	AK016546 Mus muscu
2	1669	86.3	5517	11	AK029788	AK029788 Mus muscu
3	1415	73.2	1042	12	BM557438	BM557438 AGENCOURT
4	1325.5	68.6	910	13	BO675260	BO675260 AGENCOURT
5	1297	67.1	957	13	BO675006	BO675006 AGENCOURT
6	1237.5	64.0	1041	12	BI084090	BI084090 AGENCOURT
7	1232	63.7	746	14	BY751026	BY751026 BY751026
8	1205	62.3	968	13	BO671635	BO671635 AGENCOURT
9	1204.5	62.3	3327	11	AK050021	AK050021 Mus muscu
10	1204.5	62.3	3376	11	AK050023	AK050023 Mus muscu
11	1204.5	62.3	3457	11	AK078301	AK078301 Mus muscu
12	1122	58.0	1082	13	BU239476	BU239476 60332338
13	1117	57.8	674	10	BE888655	BE888655 601513061
14	1108.5	57.3	1090	12	BM910838	BM910838 AGENCOURT
15	1099	56.9	639	10	BG165205	BG165205 602344074
16	1089	56.3	753	9	AL040398	AL040398 DXFZ4334A
17	1057	54.7	658	10	BE388695	BE388695 601286345
18	991	51.3	831	12	BS974324	BS974324 602844075
19	976.5	50.5	853	10	BF694553	BF694553 602081767
20	970.5	50.2	898	13	BU908783	BU908783 AGENCOURT
21	951.5	49.2	603	13	BU483718	BU483718 603470021
22	949	49.1	789	10	BG709118	BG709118 602675382
23	937	48.5	982	13	BO678015	BO678015 AGENCOURT
24	922	47.7	897	13	BU191638	BU191638 AGENCOURT
25	919	47.5	695	10	BG715866	BG715866 602675282
26	916.5	47.4	853	12	BI223892	BI223892 602941035
27	905	46.8	820	13	BU522948	BU522948 AGENCOURT
28	902.5	46.7	845	13	BU216915	BU216915 603109373
29	884.5	45.9	896	12	BS921463	BS921463 602824365
30	868	44.9	788	10	BI616397	BI616397 602338360
31	866	44.8	871	13	BU196183	BU196183 AGENCOURT
32	863	44.6	821	12	BI690774	BI690774 603314392
33	833	43.1	931	13	BU235022	BU235022 603409569
34	826	42.7	479	12	BI135177	BI135177 UI-M-BH3
35	788.5	40.8	1021	12	BS974587	BS974587 602844820
36	787	40.7	687	14	CB851806	CB851806 UI-CP-FW0
37	765	39.6	713	14	CA321029	CA321029 UI-M-FW0
38	762	39.4	432	13	BY520490	BY520490 BY520490
39	762	39.4	957	13	BU234147	BU234147 603412033
40	756	39.1	741	14	CB595102	CB595102 AGENCOURT
41	753	39.0	551	9	AL643656	AL643656 AL643656
42	746	38.6	735	13	BU447277	BU447277 603766685
43	738	38.2	1023	13	BY548432	BY548432 BY548432
44	736	38.1	428	13	BO681703	BO681703 AGENCOURT
45	735	38.0	1080	10	BF203783	BF203783 601868764

ALIGNMENTS

RESULT 1
LOCUS AK016546
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length cDNA
library, clone:4932434f09 product:DIPEPTIDYL PEPTIDASE 8 homolog
[Homo sapiens], full insert sequence.
ACCESSION AK016546
VERSION AK016546.1 GI:12855334
KEYWORDS
SOURCE HTCC, CAP trapper.
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE	1	Carninci, P. and Hayashizaki, Y.
AUTHORS		High-efficiency full-length cDNA cloning
TITLE		Meth. Enzymol. 303, 19-44 (1999)
JOURNAL		99279253
PUBMED		10349636
REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
AUTHORS		Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to
JOURNAL		prepare full-length cDNA libraries for rapid discovery of new genes
PUBMED		Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
AUTHORS		Kono, H., Akiyama, U., Nishi, K., Katsunai, T., Teshiro, H., Itoh, M.,
TITLE		Sumi, N., Ishii, Y., Nakamura, S., Hazada, M., Nishino, T., Harada, A.,
JOURNAL		Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K.,
PUBMED		Fujiwaka, S., Inoue, K., Togawa, K., Izawa, M., Ohara, E., Watahiki, M.,
REFERENCE	4	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
AUTHORS		Okazaki, Y., Muramatsu, M., Inoue, Y., Kir, A. and Hayashizaki, Y.
TITLE		RIKEN integrated sequence analysis (RISA) system-384 format
JOURNAL		sequencing pipeline with 384 multicapillary sequencer
PUBMED		Genome Res. 10 (11), 1757-1771 (2000)
REFERENCE	5	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
AUTHORS		Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, D., Fukuda, S.,
TITLE		Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, A.,
JOURNAL		Salto, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Salto, R.,
PUBMED		Kadota, K., Matsuda, H., Ashburner, M., Batalow, S., Casavant, T.,
REFERENCE	6	Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiya, H.,
AUTHORS		Kushl, P., Lewis, S., Matsuo, Y., Nkaido, I., Pesole, G.,
TITLE		Quackenbush, J., Schriml, L.M., Stambli, F., Suzuki, R., Tomita, M.,
JOURNAL		Wagner, L., Washio, T., Sakai, K., Okido, T., Furum, M., Aono, H.,
PUBMED		Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
REFERENCE	7	Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
AUTHORS		Platicher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
TITLE		Hochmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
JOURNAL		Marchionni, L., Mashima, J., Mazzarelli, J., Momberts, P., Nordone, P.,
PUBMED		Ring, B., Rindwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
REFERENCE	8	Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
AUTHORS		Toyo-Oka, K., Wang, K.H., Wetz, C., Whitaker, C., Wilming, L.,
TITLE		Wyszaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohitsuki, S.
JOURNAL		and Hayashizaki, Y.
PUBMED		Functional annotation of a full-length mouse cDNA collection
REFERENCE	9	Nature 409 (6821), 685-690 (2001)
AUTHORS		21085660
TITLE		11217851
JOURNAL		The FANTOM Consortium and the RIKEN Genome Exploration Research
PUBMED		Group Phase I & II Team.
REFERENCE	10	Analysis of the mouse transcriptome based on functional annotation
AUTHORS		of 60,770 full-length cDNAs
TITLE		Nature 420, 563-573 (2002)
JOURNAL		6 (bases 1 to 3143)
PUBMED		Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
REFERENCE	11	Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
AUTHORS		Furum, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
TITLE		Hataoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
JOURNAL		Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M.,
PUBMED		Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
REFERENCE	12	Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
AUTHORS		Saito, H., Salto, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
TITLE		Shibata, K., Shibata, Y., Shinagawa, A., Shitaki, T., Sugabe, Y.,
JOURNAL		Suzuki, H., Tagami, M., Tagawa, A., Takehashi, F., Tanaka, T.,
PUBMED		Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
REFERENCE	13	Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
AUTHORS		Direct Submission
TITLE		Submitted (10-JUL-2000) Yoshitake Hayashizaki, The Institute of
JOURNAL		Physical and Chemical Research (RIKEN), Laboratory for Genome

COMMENT

Exploration Research Group, RIKEN Genomic Science Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama
Kanagawa 230-0045, Japan (E-mail:genome-res@gs.c.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken.
Genomic Sciences Center and Genome Science Laboratory in RIKEN,
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5'-GAGAGAAGGAATTCACCAAGCCTCTTGTTTTGTATTVN 3']. cDNA was
prepared by using trihalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAAGAAATTCGAGTTAATTAATAATTAATCCCCCCCCCC 3']. cDNA was cleaved
with BamHI and XhoI. cDNA of size compressed longer than 7 kb was
selected before cloning. Vector: a modified plasmid pBluescript KS(+) after
bulk excision from Lambda PUC I. Cloning sites, 5' end: SalI, 3'
end: BamHI. Host: DHIOB.

FEATURES

Source

1..3143

organism="Mus musculus"

molecule_type="mRNA"

strain="C57BL/6J"

db_xref="FANTOM DB:4932434F09"

db_xref="MG1:1907554"

db_xref="taxon:10090"

clone="4932434F09"

sex="male"

tissue_type="testis"

clone_lib="RIKEN full-length enriched mouse cDNA library"

dev_stage="adult"

327..3005

/note="unnamed protein product; Dipeptidyl peptidase 8
homolog (Homo sapiens) (SPTP1|Q9HBM5, evidence: FASTY,
95.7%ID, 100%length, match=2649)

putative"

/codon_start=1

/protein_id="BAB30295.2"

/db_xref="MG1:26382128"

/db_xref="WGI:1921638"

/translation="MKIPSGRCNMAMAFETDGLVEIFFTACEBGNGSSODRPKTI
FYERSWSQLKKLLADTPRGKHGYNNAKAPHOFMYKRIDPRPHSDRYALAMS
RENTLTFSEIPIKTINRAAVLMISMKPLIDFOATLDGYMSREELRKRIGITV
AADVPYSGTFLFGSGIVPHIKDGPGHGFQOPLPRLVETSCNIMDPCLCG
PDMAITFNNDIMISLVTRERIRITYNNEILANMEDRSAGVAFFVCEEDRA
VMKCPAERTPPGGKRTILTEENDESVEITIHVS PMLETRADSFRYPKDTGNA
VTXKMSIYVDNAGGIIDVIDKEVLVDPPELTIFEGVEYIRACHTREGKANSILL
QTHLQVLISPELFEVEDANDROKLISVFPSVPLITIEETDIWINHDIRP
POHQHEIEFIETFASECKTGFRMLYKITSLSKSKYRSSGGLPADDFDCPIKEEL
TQSEMENRHGSNIWNDEARKLYVEGKMDSPLEHLVVTVANGEVVRILTDR
HSCSLSRHDCFISKSYNQKNPKCVSLYLSPEDDPVAKTEFPAATLIDSAGPLE
TPREISFESTGTFTLYGMLYKPHDLOPKKPYTLVETIYGGOVOLVNRFGVKGP
LNTLASLGVVVVINDRGSCHRGKFKREGAFKKMCOILETDPOVGLGYLAASDYPT
DRGTHGHSYGGTSLMALMQSNDIFRVAILAAPVTIMTFYTGTGTERMCHPDDN
GYLGSVAAQAQEKFPSPENRLLHLHGFDLENHFATSTILLSFLVAPKRPYLQTYI
ERSISVPSGEHYELDLAHLYLOENIGSRIALAKVI"

BASE COUNT 6 894 a

ORIGIN

688 c 773 g 788 t

Alignment Scores:

Pred. No.: 1.23e-190 Length: 3143

Score: 1756.50 Matches: 343

Percent Similarity: 65.67% Conservative: 7

Best Local Similarity: 64.35% Mismatches: 10

Query Match: 90.87% Indels: 173

DB: 11 Gaps: 1

US-10-070-464-7 (1-360) x AK016546 (1-3143)

xy 1 GluGlutAspAlaArgSerAlaGlyValAlaThrPrheValLeuGlnGluInuPheAspArg 200

Db 1086 GAAAGGATCCAGATCAAGCTGGTGGCCACCTTTGTTCTTCAAGAAATTTTGAACA 1145
 Qy 21 TysSerGlyTyrTrpCysProLysAlaGluThrProSerGlyGlyLysIleLeu 40
 Db 1146 TACTCTGGCTACTGGTGGTGTCCCAAGGGAAGAACTCTAGTGTGTAATAATCTT 1205
 Qy 41 ArgGluLeuTyrGluGlnAsnAspGluSerGluValGluIleLeuHisValThrSerPro 60
 Db 1206 AAGAAATCTCTATGAAGAAATATGATGATCTGAGGTGAGATTAATTCATGTCCTCC 1265
 Qy 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
 Db 1266 ATGTTGGAAACAAGAGGACAGATTCCTTTCGTTATCCAAAACAGCACCAACCA 1325
 Qy 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle 98
 Db 1326 AAGGTCATCTTCAAGATGTCCGAGATTTGTTGATGCTGCGAGAGGATTTAAGATGC 1385
 Qy 98 ----- 98
 Db 1386 ATGATATAAGAACTGGATTCMACTTCGAGATTCGTGTTGAGGAGTTGAATATATTGCC 1445
 Qy 98 ----- 98
 Db 1446 AAGCCGGATGATCTCCAGAGGAAACATGCTGGTCCATCTCTACTAGACCGTTCCAG 1505
 Qy 98 ----- 98
 Db 1506 ACTCAGCTGAGATAGTCTGTATCTCCCTGAGTTATTCATCCAGTAGAAGATGATGCC 1565
 Qy 98 ----- 98
 Db 1566 ATGACAGACAGAGACTATAGAGTCACTTCTGACTGTGACACCACTGATCATCTAT 1625
 Qy 98 ----- 98
 Db 1626 GAAAGAAACAAGACATCTGATTAATAATCCAGATATTTTTCATGTTTCTCAACT 1685
 Qy 98 ----- 98
 Db 1686 CATGAAGATGAATAGATTATTTTGGCTCTGATGACAAACAGTTTTCGTATCTG 1745
 Qy 98 ----- 98
 Db 1746 TATAAATCAATCCATTTTAAAGAGCAAAATATAACGGTCCAGTGTGACTACT 1805
 Qy 98 ----- 98
 Db 1806 GCCCCAAGTATTTCAAGTGTCTTATCAAAAGAAATTAACAATTACAGTGTGAATGG 1865
 Qy 99 -----ValAspGluValArgArgLeuValTyr 107
 Db 1866 GAAGTACTTGGCCGGCATGATCTAATATCTGGTTGATGAAGCCAGAAAGCTGTGTAC 1925
 Qy 108 PheGluGlyThrLysAspSerProLeuGlnHisIleuTyrValValSerTyrValAsn 127
 Db 1926 TTGGAAGCACCAAGACTCTCTTGTGAGCATCACTGTAGAGTACAGATTAATCAAC 1985
 Qy 128 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
 Db 1986 CCTGGAGAAGTGTGAGGCTGATGACCTGTGCTACTCACTCTGCTGCTGACCGCG 2045
 Qy 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
 Db 2046 CATGTGTACTTCTCATATAGTAAGTACAGCAACAGAAAGATCCACATGCTGTGCTC 2105
 Qy 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGlnPheThrAlaThr 187
 Db 2106 TACAAACTCTCAAGCTCTGAGAGTACCAAGTTCATTAACCAAGAAATTTTGGCCACC 2165
 Qy 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 207

Db 2166 ATTTGGATTACGACAGTCTCTTCTGACTACACCCCTCCGAAATTTTCTTTGAA 2225
 Qy 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPheHisAspLeuGlnProGly 227
 Db 2226 AGTACTACTGATGATTTACACTGATAGGAATGTTGTAAGCCCTCAGACTACACTGGA 2285
 Qy 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
 Db 2286 AAGAAATACCCCACTGTGTTATTCATATATGTGTGTGCCCAAGTGCAGCTGTGAACAAT 2345
 Qy 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 Db 2346 CGGTTTAAAGAGTCAAGATTTCCGCTGAACACCTGCGCTCTCTGAGTTATGTGGTT 2405
 Qy 268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
 Db 2406 GTGGTGATAGACACAGGGGATCTCTGACCGAGACTTAATTTGAAGCCCTTTAA 2465
 Qy 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
 Db 2466 TATAAAATGGGTCAATATGAATCGATGATCAAGTGAAGGAACTCCAGTACCTAGCATCT 2525
 Qy 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyTyr 327
 Db 2526 CAGTATGACTTATTGACTGATGAGTGGGATCCAGCTGAGCTGATGATGATGATGATG 2585
 Qy 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
 Db 2586 CTCTGCTGATGATGATTAATGAGAGTGGATATCTTCGGGTGCTATGCTGTGGGCC 2645
 Qy 348 ProValThrLeuThrIlePheTyrAspThrGlyTyrThr 360
 Db 2646 CCAGTCACCCGTGTGATCTTCTATGATACAGATACAG 2684
 Qy 360 ----- 360
 Db 2684 ----- 2684
 RESULT 2
 AK029788
 LOCUS
 DEFINITION
 Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:493056OC15 Product:CD15EPRIIDYL PEPTIDASE 8 homolog [Homo sapiens], full insert sequence.
 AK029788
 VERSION
 AK029788.1 GI:26081520
 KEYWORDS
 HTG; CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS
 Carninci, P. and Hayashizaki, Y.
 TITLE
 High-efficiency full-length cDNA cloning
 JOURNAL
 Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE
 99279253
 PUBMED
 10349636
 REFERENCE
 AUTHORS
 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL
 Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE
 20499374
 PUBMED
 11042159
 REFERENCE
 AUTHORS
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiwagi, K., Fujimake, S., Inoue, K., Togawa, Y., Iwawa, K., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuzaki, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 TITLE
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillarary sequencer
 JOURNAL
 Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE
 20530913

PUBMED
REFERENCE
AUTHORS
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Komuro, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kado, K., Matsuda, H., Ashburner, M., Balas, S., Casavant, T.,
Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nishikido, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Staudt, F., Suzuki, R., Tomita, M.,
Wagner, L., Mashima, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Balarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bona, M.P., Brownstein, M.T., Bulc, C., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P.,
Rind, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyo-Oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420 563-573 (2002)
6 (bases 1 to 5517)
TITLE
JOURNAL
REFERENCE
AUTHORS
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420 563-573 (2002)
6 (bases 1 to 5517)
TITLE
JOURNAL
REFERENCE
AUTHORS
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://phantom.gsc.riken.go.jp/
Location/Qualifiers
1. 5517
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM:DB:4930560C15"
/db_xref="taxon:10090"
/clone="4930560C15"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
2221. 4837
misc_feature

/note="DIPPEPTIDYL PEPTIDASE 8 homolog (Homo sapiens)
(SPTR|Q9HBM5, evidence: FASTA, 95.7%ID, 100%length,
match=2649)
putative"
BASE COUNT 1514 a 1224 c 1230 g 1549 t
ORIGIN
Alignment Scores:
Pred. No.: 3,556-180 Length: 5517
Score: 1669.00 Matches: 332
Percent Similarity: 66.67% Conservative: 10
Best Local Similarity: 64.72% Mismatches: 18
Query Match: 86.34% Indels: 133
DB: 11 Gaps: 4
US-10-070-464-7 (1-360) x AK029788 (1-5517)
QY 1 GUGUUAAPALAAAGSerLaglyValAlaThrPheValLeuGluGluPheAspArg 20
DB 2980 GAAGAGGATCCAGATCAAGCTGCTGTCACCTTGTCTTCAAGAAATTGACG 3039
QY 21 TYSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyLysIleLeu 40
DB 3040 TACTCTGGCTACTGGTGGTGTCTCCCAAGCGAAGAACTCCTAGTGTGTAATTTCT 3099
QY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleValThrSerPro 60
DB 3100 AAGATTCCTATGAAGAAATATGATGATGAGGAGATTAATCATATGCTCCGCC 3159
QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLys----- 74
DB 3160 ATGTTGAAACAAAGAGGGCAGATTCCTTTCGTTATCCAAACAGATTATGATGTCAT 3219
QY 75 -----ThrGlyThrAla----- 78
DB 3220 AGATAAGAACTGTTCAACCTTGGAGATTCGTTGAGGAGTTGAATATATGCGAG 3279
QY 78 ----- 78
DB 3280 AGCCGGATGACTCCAGAGGAAACATCGTCTTCATCTAGACCGTTCCAGAC 3339
QY 79 AsnPro-LysValAlaThrPheLysMetSerGluIleMetIleAspAlaGlu----- 94
DB 3340 TCACCTGCAATGATGTTCTGATCTCCCTGATGATTATTCATCCAGTAAAGATGATGCCAT 3399
QY 94 ----- 94
DB 3400 GGACAGACAGAGACTTATAGAGTCAGTTCCTGACTGTGACACACATCATCTATGA 3459
QY 94 ----- 94
DB 3460 AGAAACAGACAGATCGATTAATATCCAGATATTTTCATGTTTTCCTCAAACTCA 3519
QY 94 ----- 94
DB 3520 TGAAGATGAATGAGTTATTTTTCCTGCTGAAAGCAAAACAGTTTCGTATCTGTA 3579
QY 94 ----- 94
DB 3580 TAAATCATCTCATTTTAAAGAGAGCAATATAACGTCAGTGTGACTACTGC 3639
QY 94 ----- 94
DB 3640 CCCAAGTATTTCAGTCTCTATCAAGAGAAATAACAATTACCAAGTGTGTAATGGGA 3699
QY 95 -----GlyArg-----IleIleValAspGluValArgSerValTyrPhe 108
DB 3700 AGNACTTGCCCGGCAATGATCAATATCTGGGTGATGAAGCAAGAAAGCTGTACTT 3759
QY 108 eGluGlyThrLysAspSerProLeuGluIleHisIleValTyrValSerTyrValAsnPr 128
DB 3760 TGAAGCACCAAAAGACTCTCTTGGAGCATCACTGTAGCAAGTATGCAAAACC 3819

QY	128	IGL1GL1Val1ThnrArgLeuThrAspArgGlyTYrSerHisSerCysCysIleSerGlnH1	148
Db	3820	TGGAGAAATGGTGAAGCTGACTGAACCGTGGCTACTCACTCCGCTCCAGCCGCA	38797
QY	148	SCysAspPhePheIleSerLySTYrSerAsnGlnLysAsnProHisCysVal1SerLeuTY	168
Db	3880	TTGTGACTTCTTCTAATGAATGAATGACGCAACCAAGAAATGCACCTGTGTCTCTCTA	39393
QY	168	rlYsLeuSerSerProGluAspAspProThrCylAsyThrlYsGluPheTTPAlaThr11	188
Db	3940	CAAACTCTCAAGCTCTGTGAGATGACCCAGTTCATATAAACAAAGGAATTTTGGCCACAT	39993
QY	188	eLeuAspSer1aGlyProLeuProAspTYrThrProProGluIlePheSerPheGluSe	208
Db	4000	TTTGGATTTCACAGAGTCTCTCTCTGACTACACCCCTCCAGAAATTTTCTTTGAAAG	40558
QY	208	rThrThrGlyPheThrLeuTYrGlyMetLeuTYrLysProHisAspLeuGlnProGlyLY	228
Db	4060	TACTACTGATTTTACACTGATGTGAGATGTGTATTAAGCTCATGACCTACCAACCTGAAA	41119
QY	228	slyeTYrProThrValIleuPheIleTYrGlyGlyProGlnValGlnLeuValAsnAsnAr	248
Db	4120	GAATATCCCACTGTGTATTATCATATATGTTGGTCCAGAGGTCAGCTGTGGAAACATCG	41797
QY	248	gPheIleGlyValIlySTYrPheArgLeuAsnThrLeuAlaSerLeuGlyTYrValValVa	268
Db	4180	GTTTAAAGAGATCAAGTATTTTCCGCTGMAACACCTGGCTCCCTGGGTATGTGTGTGT	42339
QY	268	lVal1IleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyValaPheLysTY	288
Db	4240	GGTATATGACACACAGGGGATCTCTGCACCGAGGACTTAAATTTGAAGGCGCTTTAAATA	42999
QY	288	rLYsMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTYrLeuAlaSerAr	308
Db	4300	TAAATGGGTCAATAGAAATCGATGATCAAGTGGAGAGACTCCAGTACTCAGATCTCA	43558
QY	308	gTYrAspPheIleAspLeuAspArgValGly1IleHisGlyTrpSerTYrGlyGlyTYrLe	328
Db	4360	GTAAGACTTCATTACTTGATCGAGTGGGATCCACCGCTGGTCTCTTGTGTGGCTACTCT	44191
QY	328	uSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaPr	348
Db	4420	CTCCCTGATGSCATTAAATGACAGAGGTCCGATATCTTCCGGGTGGCTATTGCTGGGCCCC	44797
QY	348	oValThrLeuTrpIlePheTYrAspThrGlyTYrThr 360	
Db	4480	AGTCAACCTGTGTGATCTCTATGATACAGATACAG 4516	
RESULT 3			
BM557438			
LOCUS	BM557438		
DEFINITION	AGNCOCURT 6578992 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5466759		
ACCESSION	5', mRNA sequence.		
VERSION	BM557438		
KEYWORDS	BM557438.1 GI:18799430		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
ORGANISM	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
ORGANISM	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 1042)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: DCTD/DTF cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.nih.gov		

FEATURES		Plate: LICM1967 row: k column: 16	
source		High quality sequence stop: 697.	
1..1042		Location/Qualifiers	
/organism="Homo sapiens"			
/mol_type="rRNA"			
/db_xref="taxon:9606"			
/clone="IMAGE:5466759"			
/tissue_type="amelanotic melanoma, cell line"			
/lab_host="DH10B (phage-resistant)"			
/clone_id="NH MGC 41"			
/note="Organ: skin; Vector: pORF7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."			
BASE COUNT	279 a	223 c	234 g 303 t 3 others
ORIGIN			
Alignment Scores:			
Pred. No.:	4.97e-152	Length:	1042
Score:	1415.00	Matches:	262
Percent Similarity:	99.24%	Conservative:	0
Best Local Similarity:	99.24%	Mismatches:	2
Query Match:	73.20%	Indels:	0
Ds:	12	Gaps:	0
US-10-070-464-7 (1-360) x BM557438 (1-1042)			
Qy	97	11e11eValaSpGlValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu	116
Db	28	ATCCAGTTGATGANGATCAGAGGCTGGTATATTTTGAAGCACCAAGATCCCTTTA	87
Qy	117	GlulHisLeuValValSerTyrValaenProGlyGluValThrArgLeuThrAsp	136
Db	88	GAGCATCACTGTAGTACTGATCAGTTACGTAATCTCGAGAGGTGACCAAGCTGACTGAC	147
Qy	137	ArgGlyTyrSerHisSerCysCysGlnSerGlnHisCysAspPheHelleSerLysTyr	156
Db	148	CGTGCGTACTACATCTTCTGCGATCGATCGACCTGCACTGCACTTCTTTAAAGTAAGTAT	207
Qy	157	SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp	176
Db	208	AGTAAACCGAAGAAATCCACACGCTGTGTCTTTCCTTCAAGCTATCAAGTCTGAAGTGAAC	267
Qy	177	ProThrCysLysThrLysGluPheThrAlaThrLleLeuAspSerAlaGlyProLeuPro	196
Db	268	CCAACTTGCAAAACAAAGAAATTTGGGCGCACCATTTGGATTGCAAGCTCCTCTTCCCT	327
Qy	197	AspTyrThrProProGlnLlIephesePheGluSerThrThGlyPheThrLeuTyrGly	216
Db	328	GACATATCTCTCCAGAAATTTCTCTTTTGAAGTACTACTGAGATTTCATTGTATGAGG	387
Qy	237	MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle	236
Db	388	ATGCTCTCAAGACCCATCATGATCTACAGCTGGAAAGAAATATCTCATCTGCTGTTCATTA	447
Qy	237	TyrGlyGlyProGlnValGlnLeuValAsnAsnAsgPheLysGlyValLysTyrPheArg	256
Db	448	TATGGTGTCTCCAGAGTCCAGTGTGTAATATCGGTTTAAAGGAGTCAAGTATTTCCCG	507
Qy	257	LeuAsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgLysSerCys	276
Db	508	TTGATATACCTTACGCTCTCTAGGTTATGTGTGTATGATATGACACACAGGGATCTGT	567
Qy	277	HisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAsp	286
Db	568	CACCGAGGCTTAAATTTGAAGCGCCCTTTTAAATATTAATGGCTCAATATGAAATTTGAC	627
Qy	297	AspGlnValGluGlyLeuGlnTyrLeuLysSerArgTyrAspPheIleAspLeuAspArg	316

Db 628 GATCGAGTGGAGAGACTCCAAATATCTAGCTTCTCCATATGATTTCTTACCTTAGATCGT 687
 QY 317 VALGILYIIEHISGLYTPserTyrglygIYrLeuSerLeuMeAlaIeuMeGlnArg 336
 Db 688 GGGGGATCCACGGCTGCTCTATGAGATACCTCTCCCTGATGGCATTAATGACGAG 747
 QY 337 SerAspGLeuPheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyraP 356
 Db 748 TAGAGTATCTTCAGGGGTTCATTCGTCTGGGGCCCGACCTCTGTGATCTTCTATGAT 807
 QY 357 ThrGlyTyTrh 360
 Db 808 ACGAGTACACG 819
 RESULT 4
 LOCUS B0675260 910 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT 8354972 NIH_MGC_102 Homo sapiens CDNA clone IMAGE:6275478
 ACCESSION B0675260
 VERSION B0675260.1 GI:21786094
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 910)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LNCM2458 row: 1 column: 07
 High quality sequence stop: 618.
 Location/Qualifiers
 source 1..910
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6275478"
 /tissue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 102"
 /note="Organ: salivary gland; Vector: pOT81; Site_1: XhoI;
 Site_2: EcoRI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGAG(G). Library constructed
 by Ling Hong in the laboratory of Gerald M. Rubin
 (University of California, Berkeley) using ZAP-CDNA
 synthesis kit (Stratagene) and Superscript II RT (Life
 Technologies). Note: this is a NIH-MGC library."
 BASE COUNT 248 a 186 c 212 g 262 t 2 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 8.28e-142 Length: 910
 Score: 1325.50 Matches: 254
 Percent Similarity: 96.24% Conservative: 2
 Best Local Similarity: 95.49% Mismatches: 6
 Query Match: 68.57% Indels: 4
 DB: 13 Gaps: 1
 US-10-070-464-7 (1-360) x B0675260 (1-910)
 QY 97 IIEIIEVALASPGIUALARGATGLeuValTYrPheGluGlyThryLysAspSerProLeu 116

Db 116 ATCCAGTGTGAGAGAGTCAAGAGCTGTGATATTTTGAAGACCAAGACTCCCTTTA 175
 QY 117 GIuHISLLeuTYrValValSerTYrValAsnProGlyGluValThrArgLeuThraP 136
 Db 176 GAGCATCACTCTAGCTAGTACGTTACGTAATCTCGAGAGAGCTGACAAAGCTGACTGAC 235
 QY 137 ArgGlyTYrSerHisSerCysAlleSerGlnHisCysAspPhePheIleSerLYr 156
 Db 236 CGTGCTACTTCACATTCCTTGCTGCATCACTGACACACTGACCTCTTTAATGATGAT 295
 QY 157 SerAsnGlnLeuAsnProHisCysValSerLeuTYrLYrLeuSerSerProGluAspAsp 176
 Db 296 AGTAACCAAGAAATCCACACTGTGTGCTCTTCAACACTATCAAGTCTCAAGATGAC 355
 QY 177 ProThrCysLYrThryLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 196
 Db 356 CCAACTGGCAAAACAAGAAATTTGGGCGCACATTTGGATTGACAGAGTCTCTTCCT 415
 QY 197 AspTYrThrProProGluIlePheSerPheGlnSerThrThryGlyPheThryLYr 216
 Db 416 GACTATACTCTCCAGAAATTTTCTTTTGAAGACTACTGATTTACATTTGATGGG 475
 QY 217 MetLeuTYrLYrProHisAspLeuGlnProGlyLYrLYrTYrProThrValLeuPheIle 236
 Db 476 ATGCTCTACAAGCTCATGATCTACAGCTGGAAGAAATATCTGCTGCTTCATA 535
 QY 237 TYrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLYrGlyValLYrTYrPheArg 256
 Db 536 TATGGGAGTCCACAGGTGAGTGGATTAATCCGTTTAAAGAGACTCAAGTATTTCCGC 595
 QY 257 LeuAspThrLeuAlaSerLeuGlyTYrValValValIleAspAsnArgLYrSerCys 276
 Db 596 TTGAATACCTTACCTCTTACGATATGATGTGATGATGATGACAAAGGGATCTGT 655
 QY 277 HISATGGLYLeuLYrPheGluGlyAlaPheLYrTYrLYrMetGlyGlnIleGluIleAsp 296
 Db 656 CACCGAGGGCTTAATATTGAAGCGCTTTAATATTAATGGCTCMAATGAGAAATGAC 715
 QY 297 AspGlnValGluGlyLeuGlnTYrLeuAlaSerArgTYrAspPheIleAspLeuAspArg 316
 Db 716 GATCAGGTGGAAGAGCTCCAAATATCTAGCTTCGATATGATTTCACTTACCTTAGATCGT 775
 QY 317 VALGILYIIEHISGLYTPserTyrglygIYrLeuSerLeuMeAlaIeuMeGlnArg 336
 Db 776 GGGGGATCCACGGCTGCTCTATGAGATACCTCTCCCTGATGGCATTTAATGACGAA 835
 QY 336 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrp---Ile-Phe 354
 Db 836 GTCAGATTATCTTCAAGTTGCTATTCGTGGGGCCCAANTCAATTCTGGGGGATCTTTC 895
 QY 355 TYrAspThryGly 358
 Db 896 TATGATACAGGG 907
 RESULT 5
 LOCUS B0675006 957 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT 8303592 NIH_MGC_102 Homo sapiens CDNA clone IMAGE:6274622
 ACCESSION B0675006
 VERSION B0675006.1 GI:21785840
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 957)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LNCM2456 row: h column: 15
 High quality sequence start: 4
 High quality sequence stop: 624.
 Location/Qualifiers

FEATURES

source

1..957
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6274622"
 /tissue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_102"
 /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGG(G). Library constructed
 by Ling Hong in the laboratory of Gerald M. Rubin
 (University of California, Berkeley) using ZAP-cDNA
 synthesis kit (Stratagene) and Superscript II RT (Life
 Technologies). Note: this is a NIH_MGC library."
 BASE COUNT 257 a 206 c 228 g 266 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.72e-138 Length: 957
 Score: 1297.00 Matches: 253
 Percent Similarity: 95.15% Conservative: 2
 Best Local Similarity: 94.40% Mismatches: 8
 Query Match: 67.10% Indels: 5
 DB: 13 Gaps: 0

US-10-070-464-7 (1-360) x BQ675006 (1-957)

QY 97 lleleValaspGluValARgArGleuValTYrPheGluGlyThrlYsaPseProleu 116
 Db 129 ATCCAAAGTtGATGAAGTtCAAGAGGCTGATATTTTGAAGGACCAAGACCTCCCTTTA 188
 QY 117 GluHSHSLeuTYrValValSerTYrValAsnProGlyGluValThraArgLeuThraP 136
 Db 189 GAGCATCACTGAGTACTGATCACTGTAATCTCGAAGGTGACAAAGGCTGATGAC 248
 QY 137 ArgGlyTYrSerHisSerCysIleSerGlnHisCysAspPhePheIleSerTYr 156
 Db 249 CCGTGCTACTCACTTCTTGCTGACATCACTGACACTGCTCTTTATTAAGTAAGTAT 308
 QY 157 SerAsnGlnLYsaAsnProHisCysValSerLeuTYrLYsaSerSerProGluAspAP 176
 Db 308 AGTAACCAAGAAATCAACACTGTGTCTCTTCAAGCTTATCAAGTCTCGAAGAGAC 368
 QY 177 ProThrCysLYsaThrlYsGluPheTrpAlaTrnIleLeuAspSerIaGlyProLeuPro 196
 Db 369 CCAACTGCAGAAACAAAGAAATTTTGGCCACATTTTGATTCAGCAGTCTCTTCT 428
 QY 197 AspyTYrThrProGluIlePheSerPheGluSerThrlYsGlyPheThrlYsGly 216
 Db 439 GACTATACCTCCCGAAGAAATTTCTCTTTGAAGACTACGATTTACATTTGATGAG 488
 QY 217 MetLeuTYrLYsaProHisAspLeuGlnProGlyLYsaLYsaTYrProThrValLeuPhe 236
 Db 489 ATGCTCTACAGCCTCATATCTACAGCCTCGAAGAAATATCTTCTGCTTTCAATA 548
 QY 237 TYrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLYsaLYsaLYsTYrPheArg 256
 Db 549 TATGTTGTGCTCAGGTGAGTGTGATTAATCGGTTTAAAGAGACTCAAGATATTTCCGC 608
 QY 257 LeuAsnThr-LeuAlaSerLeuGlyTYrValValValIleAsp-AsnArgGlySerC 276

Db 609 TTGAATACCCCTTAGGCTCTTAGTATGTGTGTAGTATGATGACAAACAGGGATCCT 668
 QY 276 YHILARGGLYLeuLYsaPheGluGlyAla-PheLYsTYrLYsaMet-GlyGlnIleGluI 295
 Db 669 GTCAACGAGGGCTTAAATTTGAAGGCGCCCTTTTATATATAAATGGGTCAAATAGAAAT 728
 QY 285 eAspAspGlnValGluGlyLeuGlnTYrLeuAlaSerArgTYrAspPheIleAspLeu 315
 Db 729 TCACGATGAGGGGAGAGAGCTCCAAATATCTAGCTTCTCGATATGATTTATTGACTTAGA 788
 QY 315 PARValGlyIleHISGlyTYrSerTYrGlyGlyTYrLeuSerLeuMetAlaLeuMetG 335
 Db 789 TCGTGTGGGCATCCACGCGTGTCTATGAGAGATCCCTCCCTGAAAGCATTTATGCA 848
 QY 335 rArgSerAspIlePheArgValAlaIleAlaGlyAlaPProValThr-LeuTrpIlePheT 355
 Db 849 GAGGTGAGATATCTTCAGAGTGTCTATGTCTGGGGGCCAGATCATCCGTTGATCTTCC 908
 QY 355 YrAspThrGlyTYr 359
 Db 909 ATGATACAGGATAC 922

RESULT 6
 BI084090
 LOCUS
 DEFINITION
 mRNA sequence.
 ACCESSION
 BI084090.1 GI:14502420
 VERSION
 BI084090.1
 KEYWORDS
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1 (bases 1 to 1041)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 AUTHORS
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 Unpublished
 COMMENT
 Contract: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LNCM1820 row: j column: 13
 High quality sequence stop: 847.
 Location/Qualifiers

FEATURES

source

1..1041
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5013996"
 /tissue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_102"
 /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGG(G). Library constructed
 by Ling Hong in the laboratory of Gerald M. Rubin
 (University of California, Berkeley) using ZAP-cDNA
 synthesis kit (Stratagene) and Superscript II RT (Life
 Technologies). Note: this is a NIH_MGC library."
 BASE COUNT 280 a 227 c 242 g 292 t

ORIGIN

Alignment Scores:
 Pred. No.: 1.41e-131 Length: 1041
 Score: 1237.50 Matches: 254

Percent Similarity: 93.77%
 Best Local Similarity: 93.04%
 Query Match: 64.02%
 DB: 12
 Gaps: 0

US-10-070-464-7 (1-360) x BI084090 (1-1041)

QY 97 IIEIIEVLAASPGLIValArgArgLeuValTyrPheGluGlyThrIlysAspSerProLeu 116
 DB 120 ATCCAGTTGATGAACTCAGAAAGGCTGGATATATTTGAAAGGCGCCAAACACTCCCTTTA 179
 QY 117 GIUHSIIEuTyrTyrValIleSerTyrValAspProGlyGluValThrArgLeuThrAsp 136
 DB 180 GAGCATCCCTGTTAGCTAGTCACTTTCGTAATCTCTGGAGAGGTGACAGGCTGACTGAC 239
 QY 137 ATGGLYTYrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerIlyr 156
 DB 240 CGTGCTACTCATTCTTGCTGTCATCAGACAGACCTGAGACTTTTATTAAGTAACTAT 299
 QY 157 SerAsnGlnLysAsnProHisCysValSerLeuTyrIlysLeuSerSerProGluAspAsp 176
 DB 300 AGTAACCAAGATCCACACTGTGTCTCTTTACAGCTATCAAGTCTCTGAAAGATGAC 359
 QY 177 ProThrCysIleSerThrIlysgluPheTyrPalaThrIleLeuAspSerAlaGlyProLeuPro 196
 DB 360 CCACTTGCAGAAACAAAGAAATTTGGCCACACTTTGGATTCAAGAGTCTCTTCT 419
 QY 197 AAPTYYrThProProGluIlePheSerPheGluSerThrThrGlyPheThrIleuTyrgly 216
 DB 420 GACTAATACCTCCCAAAATTTCTCTTTGAAAGTACTACGATTTACATTGTATGGG 479
 QY 217 MetLeuTyrIlyserProHisAspLeuGlnProGlyIlyleuTyrProThrValIleuPheIle 236
 DB 480 ATGCTCTACAGAGCTCATGATCTACAGCTCGAAAGAAATCTTACTGTCTGTTCA 539
 QY 237 TYrGlyGlyProGlnValGlnLeuValAsnAsnArgPheIlysgly--ValIlystyPhea 256
 DB 540 TATGGTGGTCTCCTAGGTGAGTGTGTAATACGCTTTAACGCGAGTCAATATTTCC 599
 QY 256 rgleuAsnThrIleuValAserIleuGlyTyrValValValIleAspAsnArgIlySerC 276
 DB 600 GCTTGAATACCCAGCTCTCTGATGTTAGTGTAGTATGAGTACCAACAGGGGATCC 659
 QY 276 YehIISrGlyLeuLysPheGluGlyAlaPheIlystyTyrIlyMetGlyGlnIleGlnIleA 296
 DB 660 GTACCGAGGCTTAATTTGAAAGGCGCTTTAA-TATCAATGGGTCAATA-GAATTTG 717
 QY 296 SPASPGLIValGluGlyLeuGln-TyrLeuAlaSerArgTyrAspPheIle-AspLeuAs 315
 DB 718 AGATCAAGTGGAGAGACTCCCAATTAATCTAGCTTCCGATATGATTTCACTTGA 777
 QY 315 PATGVALIGlyIleHisGlyTyrSerTyrGly-GlyTyrIleuSerIleu-MetAlaLeuMet 334
 DB 778 TCGTGGGATCCACAGGCTGTGTCATAGAGGAGTAATCTCTCCGAAATGGCATTAAG 837
 QY 335 GlnArgSerAsp-IlePheArgValAlaIleAla-GlyAlaProValThrLeuTyrIleP 354
 DB 838 CAGAGGTCAATACCTTCCAGAGGTCAATAGCTGGGGGCCCAAGTCACTCTGTGATCT 897
 QY 354 he-TyrAspThrGlyTyrThr 360
 DB 898 TCTTATGATACAGGATACAG 918

RESULT 7
 BY751026 746 bp mRNA linear EST 17-DEC-2002
 LOCUS BY751026
 DEFINITION BY751026 RIKEN full-length enriched, activated spleen Mus musculus
 ACCESSION CDNA clone F830018D07 5', mRNA sequence.
 VERSION BY751026
 KEYWORDS BY751026.1 GI:27181193
 EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 746)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Oosato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojodori, I., Balderelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Knapkin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Brusic, V., Chothia, C., Corbett, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimond, S., Gusticich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kani, A., Kawaji, H., Kawasawa, Y., Kedierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Namata, K., Okido, T., Pavan, W. J., Pettes, G., Pesole, G., Petrovsky, N., Pillai, R., Pontus, J. U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semp, C. A., Setou, M., Shimada, K., Sultana, R., Takemura, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wyszaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arikawa, T., Fukuda, S., Hara, A., Hashizume, M., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, U., Kojima, Y., Komodo, S., Komno, K., Koyu, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

TITLE JOURNAL MEDLINE PubMed

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, U., Kojima, Y., Komodo, S., Komno, K., Koyu, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 eComputer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/WRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers

1..746

/organism="Mus musculus"

/mol_type="mRNA"

/strain="NOD"

/db_xref="taxon:10090"

/clone="F830018D07"

/issue_type="activated spleen"

/clone_1ib="RIKEN full-length enriched, activated spleen"

BASE COUNT 194 a 178 c 176 g 197 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 3.62e-131 Length: 746

Score: 1232.00 Matches: 229

Percent Similarity: 97.48% Conservative: 3

Best Local Similarity: 96.22% Mismatches: 5

Query Match: 63.74% Indels: 1

DB: 14 Gaps: 0

US-10-070-464-7 (1-360) x BY751026 (1-746)

QY 124 SerTyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCys 143

DB 1 AGTTATGCAAAACCTGAGAGAGTGTGAGGCTGACTGACCGTGGCTAATCACTCCCTGC 60

QY 144 CysIleSerGlnHisCysAspPhePheIleSerIleSerAsnGlnAspProHis 163

DB 61 TGGCTACGCGGATGTGACTTCTTCATAGTAAGTACAGCAACAGAAAGATCAACAC 120

QY 164 CysValSerLeuTyrIleuSerSerProGluAspAspProThrCysIleTyrIleGlu 183

DB 121 TGTGTGTCCCTCACAACCTCAAGTCCCTGAGAGTACCCAGTTCATTAACAAAGAA 180

QY 184 PheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspPheThrProProGluIle 203

DB 181 TTGGGACCAACCTTGGATGAGAGGTCCTCTTCGACTACACCCCTCAGAAAT 240

QY 204 PheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleProHisAsp 223

DB 241 TTTCTTTGAAAGTACTGATGATTACCTGTATGAGATGTGTATTAAGCTCATGAC 300

QY 224 LeuGlnProGlyIleuSerIleProThrValLeuPheIleTyrGlyGlyProGlnValGln 243

DB 301 CTACAACTCGAAGAAATACCCCACTGTATATATATATATATATATATATATATATAT 360

QY 244 LeuValAsnAspArgPheIleGlyValIleTyrPheArgLeuAsnThrIleuAspSerIleu 263

DB 361 CTGGTGAACATCGGTTTAAAGAGTCAAGTATTTCCGCTGAAACCCCTGGCTCCCTG 420

QY 264 GlyTyrValValValValIleAspAsnArgIleSerCysHisArgGlyLeuIleuSerPheGlu 283

DB 421 GGTATGTGTGTGTGTGTATAGACAAAGGAGTCCCTGCAACGAGAGCTTAATTTGAA 480

QY 284 GlyAlaPheLeuTyrIleuSerGlyGlnIleGlnIleAspAspGlnValGlnIleLeuGln 303

DB 481 GGGGCTTTAAAT 540

QY 304 TyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSer 323

DB 541 TACCTAGCATCTCAGATGATGATCTTCACTTGGATCGAGTGGGATCCACGAGCTGATCC 600

QY 324 TyrGlyGlyTyrIleuSerLeuMetAlaLeuMetGlnArgSerAspIle-PheArgValAla 343

DB 601 TATGGTGGTACTCTCCCTGATGGCATTAATGCAAGGTCGATATCTCTTCGAGTGGG 660

QY 343 AlieAlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThr 360

DB 661 TATTGTGGGGCCCGCAGTCACTGTGATCTTCTATGATACAGATACAG 712

RESULT 8

B0671635

LOCUS B0671635 968 bp mRNA linear EST 15-JUL-2002

DEFINITION

AGENCOURT 8203887 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6255645

5', mRNA sequence.

ACCESSION

B0671635

VERSION

B0671635.1 GI:21782469

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 968)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

CONTACT: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

plate: L1CM2407 row: a column: 22

High quality sequence stop: 630.

Location/Qualifiers

1..968

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6255645"

/issue_type="epidermoid carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_1ib="NIH MGC 102"

/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."

BASE COUNT 245 a 228 c 235 g 259 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 7e-128 Length: 968

Score: 1205.00 Matches: 229

Percent Similarity: 97.46% Conservative: 1

Best Local Similarity: 97.03% Mismatches: 3

Query Match: 62.34% Indels: 3

DB: 13 Gaps: 0

US-10-070-464-7 (1-360) x B0671635 (1-968)

QY 97 IleIleValAspGluValArgArgLeuValTyrPheGluGlyThrIleAspSerProLeu 116

DB 117 ATCCAAGTGAAGTGAAGTCAAGAGCTGATATATTTGAAGGCACCAAGCTCCCTTTA 176

QY 117 GluHisIleLeuTyrAlaValSerTyrValAspProGlyGluValThrArgLeuTyrAsp 136

DB 177 GAGCATCACTGATCACTGATCAAGTTCGTAATCCCTGAGAGGTGACAAAGCTGACTGAC 226

QY 137 ArgGlyTyrSerHisSerCysIleSerGlnHisCysAspPhePheIleSerIleTyr 156

DB 237 CGGGCTACTCACTCTTGGCTGATCACTGATCACTGATCACTCTTTAATTAAGTAT 236

QY 157 SerAsnGlnLeuAsnProHisCysValSerLeuTyrIleuSerSerProGluAspAsp 176

DB 297 AGTAACCAAGAAGATCAACACTGTGTGCTCTTCAAGCTATCAAGTCCCTGAAGATGAC 356

QY 177 ProThrCysValThrIleGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuPro 196

DB 357 CCAACTGCAAAACAAAGAAATTTGGGCCACCACTTTGATTCACAGGTCTCTTCT 416

QY	197	AspGlyThrProPheGluIlePheSerPheGluSerThrThiGlyPheThrLeuTyGly	216
Db	417	GACATATCTCTCCAGAAATTTTCTCTTTGAAAGTACTACTGATTAACATTGATGG	476
QY	217	MetLeuTyIysProHisAspLeuGlnProGlyIysIysIyrProThrValIeuPheIle	236
Db	477	ATGCTCTACAAAGCCTCAATGATCTTCACGCTGGAAAGAAATATCCACTGCTGTTCATA	536
QY	237	TyrGlyIysIyrProGlnValGlnLeuValAsnAsnAargPheIysGlyValIlyTyPheArg	256
Db	537	TATGCTGCTCTCAGGTCAGTGTGTAATATCGTTTAAAGGAGTCAAGATATTTCCGC	596
QY	257	LeuAsnThrLeuAlaSerLeuGlyTyrTyrValValIleAlaAspAsnArgGlySerCys	276
Db	597	TTGAATACCTCAGCTCTCTGAGTTATGTGTGTAGATAGACAAACAGGGATCTCTGT	656
QY	277	HisArgGlyLeuIysPheGluGlyValAlaPheIysIyrIysMetGlyGlnIleGluIleAsp	296
Db	657	CACCGAGGGCTTAAATTTGAAGAGCGCCTTANAATATAAATGGGTCAATAGAAATTTGAC	716
QY	297	AspGlnValGluGlyLeuGlnTyrIleuAlaSerArgTyrAspPheIleAspLeuAspArg	316
Db	717	GATCAGGGGAGGAGCTCAATATCTACTGCTTCGATATGATTAATGACTTAAATCGG	776
QY	317	-ValGlyIleHisGlyTyrPser-Tyr-GlyGlyIyrLeuSer 329	
Db	777	TGTGGGAGCTTCCACGCGCTGGGCCCTATGAGGAGATACCTCTCC 818	
RESULT 9			
AK050021		3327 bp mRNA linear HTC 05-DEC-2002	
LOCUS			
DEFINITION		Mus musculus adult male liver tumor cDNA, RIKEN full-length	
VERSION		enriched library, clone: C790303D12 product: D19PRTIDYL PBT7IDM5 9	
KEYWORDS		homolog [Homo sapiens], full insert sequence.	
SOURCE		AK050021.1 GI:26340743	
ORGANISM		HTC, CAP trapper.	
		Mus musculus (house mouse)	
REFERENCE			
AUTHORS		Bukaryovci, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS		1	
TITLE		Carninci, P. and Hayashizaki, Y.	
JOURNAL		High-efficiency full-length cDNA cloning	
MEDLINE		Meth. Enzymol. 303, 19-44 (1999)	
PUBMED		99279253	
REFERENCE		10349636	
AUTHORS		2	
TITLE		Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,	
JOURNAL		Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.	
MEDLINE		Normalization and subtraction of cap-trapper-selected cDNAs to	
PUBMED		prepare full-length cDNA libraries for rapid discovery of new genes	
REFERENCE		Genome Res. 10 (10), 1617-1630 (2000)	
AUTHORS		3	
TITLE		Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,	
JOURNAL		Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Teshiro, H., Itoh, M.,	
MEDLINE		Sun, N., Ishii, Y., Nakamura, S., Hazama, N., Nishine, T., Harada, A.,	
PUBMED		Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,	
REFERENCE		Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,	
AUTHORS		Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,	
		Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.	
TITLE		RIKEN integrated sequence analysis (RISA) system-384-format	
JOURNAL		sequencing pipeline with 384 multicapillary sequencer	
MEDLINE		Genome Res. 10 (11), 1757-1771 (2000)	
PUBMED		20530913	
REFERENCE		11076861	
AUTHORS		4	
TITLE		Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,	
JOURNAL		Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, Y., Fukuda, S.,	
MEDLINE		Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, T.,	
PUBMED			
REFERENCE			
AUTHORS			

Salto, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Salto, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuenli, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schiraldi, L. M., Staudli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustinchik, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzaletti, J., Mommaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodrigues, I., Sakamoto, N., Saeki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittraker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S. and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

11217851

5

THE RANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 3327)

Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hoti, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Salto, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Saeki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tanaka, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-JUL-2001) Yoshhide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Science Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Science Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.

Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
Location/Qualifiers

1. .3337

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM,DB:C730003D12"
/db_xref="taxon:10090"
/clone="C730003D12"
/sex="male"
/cell_type="tumor"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
176. .2764

CDS

LOCUS	AK050023	3376 bp	mRNA	linear	HTC 05-DEC-2002
DEFINITION	Mus musculus adult male liver tumor cDNA, RIKEN full-length enriched library, clone: C730003H19 product: D1PEPTIDYL PEPTIDASE 9 homolog (Homo sapiens), full insert sequence.				
ACCESSION	AK050023				
VERSION	AK050023.1	GI:26093756			
KEYWORDS	HTC, CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Enkaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Math. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PubMed	10349636				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20493574				
PubMed	11042159				
REFERENCE	3				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Matsuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuzura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multipipillary sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				
MEDLINE	20530913				
PubMed	11076861				
REFERENCE	4				
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Iwawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakado, I., Pesole, G., Quackenbush, J., Schriml, L. M., Steinhil, F., Suzuki, R., Tomita, M., Wadner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Madorell, R., Barth, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustinich, S., Hill, D., Hotmann, K., Hume, D. A., Kamlay, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazarrelli, J., Mommaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyokawa, K., Wang, K. H., Weitz, C., Wiltaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S. and Hayashizaki, Y.				
TITLE	Functional annotation of a full-length mouse cDNA collection				
JOURNAL	Nature 409 (6821), 685-690 (2001)				
MEDLINE	21085660				
PubMed	11217851				
REFERENCE	5				
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.				
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
JOURNAL	Nature 420, 563-573 (2002)				
REFERENCE	6 (bases 1 to 3376)				
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,				

[illegible]

Db	1112	GCCTTGAGGAGAGAGAGACGAGCTCCTTACCGCTACCCAGAGACGAGCAGAAAGACCC	1171
QY	81	LysValThrPheIysMetSerGluIleMetIleAspAlaGluIleValIle-----	98
Db	1172	AAAGATTGCCTGAAGCTGGCTGAAGCTCGACGAGCATCAGGCGAAATCTGTCTCAAGC	1231
QY	98	-----	98
Db	1232	TGCGAGAAAGAACTGGTACAGCCATTGACGTCCTTTCCCAAAAGTGAATACATCGCC	1291
QY	98	-----	98
Db	1292	CGGCGTGGCTGGACACGAGGACGAGCAAAATATGCTGGGCAATGTTCTTGACCGTCCAG	1351
QY	98	-----	98
Db	1352	CAACGGCTTACAGCTTTCCTCTGCGCCCTGCTCTTTCATCCCGGCGTTGAGAGTGAG	1411
QY	98	-----	98
Db	1412	GCCGACGCGAGGAGAGCTGCCAGAGCGGTCCCAAGAAATGTCAGAGCCTTTGTGATCTAT	1471
QY	98	-----	98
Db	1472	GAAAGAGTCACCAATGTCTGGATCAAGCTCCACGACATCTTCCACCGGTTTCTCAGGCT	1531
QY	98	-----	98
Db	1532	GAGGGCAGACAGACTTTTGTTCCTTGTGTCACCAAGATCAGACATGAGTTCGACAC	1591
QY	98	-----	98
Db	1592	CTGTACAGAGGTCAAGTGGAACTTAAAAACAAGACTATGACTGACGAAACCCCTCAGC	1651
QY	98	-----	98
Db	1652	CCTACAGAGATGAGTTTAAGTCCCACTCAAGAGAGAGTCCGCTGACCAAGTGGCAG	1711
QY	99	-----ValAspGluValArgGluVal :::	106
Db	1712	TGGAGAGCTGTGAGGATAGGCTCCAAAGATCTGGGTCACAGACAGACAGACAGCTGGTG	1771
QY	107	TyrPheGluGlyThrIysAspSerProLeuGluHisIleValValValSerTyrVal	126
Db	1772	TACTTCAAGGTACAAAGAACACACCGGTGAAACATCACTCTATGATGTAGCTACAG	1831
QY	127	AsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSer	146
Db	1832	TCACGAGCGAGATCGTGCCTGCTACCAACGCTCGGCTCTTCCACAGCTGTCTCATGAGC	1891
QY	147	GlnHisCysAspPhePheIleSerIleSerIleSerIleAsnGlnIlyAsnProHisCysValSer	166
Db	1892	CAGAGCTTCGACATGTTCTGTGATGATCACTACAGACAGGTGAGACAGCCACCTGTGTACAT	1951
QY	167	LeuTyrIlyIleAspSerProGluAspAspProThrCysIlyThrIyGluPheTyrAla	186
Db	1952	GTGTCACAGCTGAGACGGCCCGCAGATATACCACTGACCAAGCAACACACGCTTTCGGCC	2011
QY	187	ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPhe	206
Db	2012	AGCATGTATGGAGGACCAATTTGCCCCCAAGACTATGTGCCCTCGAATCTTCCACATTC	2071
QY	207	GluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIlyAspProHisAspLeuGlnPro	226
Db	2072	CACACCCGTGACAGACGTGACGCTCTTACGGCATGTATCAAGACCAACACACCTTGACACT	2131
QY	227	GlyIlyIlyIysTyrProThrValIleuPheIleTyrGlyGlyProGluValGlnLeuValAsn	246
Db	2132	GGGAGGAAAGACCCCACTGCTCTTGTGTATGGGGGCCCAACAGGTGACAGTGGTGAAC	2191
QY	247	AsnArgPheIysGlyValIlyTyrPheArgLeuAsnThrIleuAlaSerLeuGlyTyrVal	266
Db	2192	AAATCTTTTAAAGGAGATAAATACCTGCGGCTAAATTAACAGGACATCTTGGGCTATAGCT	2251

Qy	26	ValValVal11leapspenaVgglYseCyShiAriXgLYleuLysPhegluGlyAlaPhe	286
Db	2252	GtGtGtGtGtATCCATGCTCGAGGCTCTCTGTCAAGTGGGACCTTCGACCTTCAGAGGGGCTTG	2111
Qy	287	LyvTyTyLysMetGlyGlnIleGlnIleaspAspGlnVAlGlnGlyLeuGlnTyTLeuAla	306
Db	2312	AAAAATCAATGGGCCACGTCGAGATTCAGAGACCAAGTCGGAAGCGCTTCGACGTACGTGGCT	2371
Qy	307	SeArgrTyAspPheIleleaspLeuAspArGValGlyTleHieglYTPseTYTgLYGly	326
Db	2372	GAGAGATGATGCGCTTCATGATGACTTCGAGCGCGATGCCCATTCATGGCTGCTGCTTACGGCGC	2431
Qy	327	TyTleSeSerleuMetAlaLeuMetGlnArSseArspIlePheArGValAlaIleAlaGly	346
Db	2432	TTCTCTCATCTCATGGGGCTTCATCCACAAAGCCACCAATGTTCACAGGATGGCATTTGGCGGC	2491
Qy	347	AlaProValThrIleuTPeIlePheTyAspThrGlyTYrThr	360
Db	2492	GCTCTCTCATCTGTGTGATGGCTTATGACACAGGGATCACGC	2533
RESULT 11			
AK078301			
LOCUS			
DEFINITION			
AK078301			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			


```

Db      1598 CCTACAGAAAGATGAGTTTAAGTCCCATCAAGAGAGGTGCGCTTGACCAAGTGGCGAG 1657
Qy      99 -----ValAspGluValArgArgLeuVal 106
Db      1658 TGGAGGCTTGTCTCAAGGATGAGCTCCAAAGATCTGGGTCAACAGACAGCAAGCTGGTG 1717
Qy      107 TyrPheGluGlyThrLysAspSerProLeuGluHisLysLeuValValSerTyrVal 126
Db      1718 TACTTCAAGGTACAAAGAGACACCGCTGGAAATCACTTATGTGTGCTGACAGAG 1777
Qy      127 AsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysLysLeu 146
Db      1778 TCAGACGGCGAGATGTGGGTGCTCAACGCTGGCTTCCCAAGCTGCTCCCAATGAGC 1837
Qy      147 GlnHisCysAspPhePheLeuSerLysTyrSerAsnGlnLysAsnProHisCysValSer 166
Db      1838 CAGAGCTTCGACATTTGTGTAGTCACTACAGCACTGACAGACAGCCCTGTGTACT 1897
Qy      167 LeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrPala 186
Db      1898 GTGTCAAGCTGAGCGGCGCCCGATGATGACCCATGCAACAGCAACAGCTTCTGGGCG 1957
Qy      187 ThrLysLeuAspSerValGlyProLeuProAspTyrThrProProGluLysPheSerPhe 206
Db      1958 AGCATGATGAGAGCGACCCCAATTGCCGCCAGACTATGTGCCGCCCTGAGATCTTCCACTTC 2017
Qy      207 GluSerThrThrGlyPheThrLysMetLeuTyrLysProHisAspLeuGlnPro 226
Db      2018 CACACCCGTCGACAGCTGAGCTTACGGCATGATCTACAGACACACCTGCAACT 2077
Qy      227 GlyLysLysTyrProThrValLeuPheLeuTyrGlyGlyProGluValGlnLeuValAsn 246
Db      2078 GGGAGGAAGACACCCCACTGTCTTGTCTATGGGGGCCCAAGGTGACAGTTGGTGAAC 2137
Qy      247 AsnAspPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrVal 266
Db      2138 AACCTCTTAAAGGGATCAAAATACCTGCGCTAAATACACTGCGCACTTGTGGCTATGCT 2197
Qy      267 ValValValAlaLeuAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGlyAlaPhe 286
Db      2198 GTGGTGGTATGATGATGCTGGGGCTCTCTCAACCGGGGCTCTCACTTCAAGGGGCGCTCG 2257
Qy      287 LysTyrLysMetCysGlnIleGluLysAspAspGlnValGluGlyLeuGlnTyrLeuAla 306
Db      2258 AAAAATCAATGCGCCAGAGGTGAGATTGAGACCAAGGTGAAAGGCTTGACGTAACGGCT 2317
Qy      307 SerArgTyrAspPheLeuLysPheLysPheValGlyIleHisGlyThrSerTyrGlyGly 326
Db      2318 GAGAAGTATGGCTTCACTTACCTTGAAGCCGAGTCCGCTCACTGCTTACGCGCGC 2377
Qy      327 TyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 346
Db      2378 TTCTCTCACTCACTGAGGCTCACTCCACCAAGCCACCAAGTTCACAGTTCACCTTGGGCG 2437
Qy      347 AlaProValThrLeuTyrPheTyrAspThrGlyTyrThr 360
Db      2438 GCTCCGTGCTGCTGATGATGCTGCTATGACACAGGCTACACG 2479

```

```

RESULT 12
BU239476      1082 bp      mRNA      linear      EST 26-NOV-2002
LOCUS      BU239476      603223338F1 CSEQCHN33 Gallus gallus cDNA clone CHEST24919 5', mRNA
DEFINITION      sequence.
ACCESSION      BU239476      GI:25485538
VERSION      BU239476.1      GI:25485538
KEYWORDS      EST.
SOURCE      Gallus gallus (chicken)
ORGANISM      Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE      1 (bases 1 to 1082)
AUTHORS      Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,

```

```

TITLE      Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
JOURNAL      A Comprehensive Collection of Chicken cDNAs
MEDLINE      Curr. Biol. 12 (22), 1965-1969 (2002)
PUBMED      22335534
COMMENT      012445342
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES
source
1..1082
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton line 151"
/db_xref="taxon:9031"
/clone="CHST24919"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHN33"
/note="Organ: liver; Vector: pBluescript II KS(+); Site_1:
Score; Site_2: Not1; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was bluntend, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."
BASE COUNT      289 a      240 c      280 g      273 t
ORIGIN
Alignment Scores:
Pred. No.: 0      3.05e-118      Length: 1082
Score: 1122.00      Matches: 217
Percent Similarity: 88.01%      Conservative: 18
Best Local Similarity: 81.27%      Mismatches: 27
Query Match: 58.04%      Indels: 5
DB: 13      Gaps: 2
US-10-070-464-7 (1-360) x BU239476 (1-1082)
Qy      97 IleIleValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 116
Db      107 ATCTATGTATGATGAAGAACCAAAATCTGCTATTTTCAAGGACGAAAGACTCTCTTTA 166
Qy      117 GlnHisLysLeuTyrValValSerTyrValAspProGlyGluValThrArgLeuThrAsp 136
Db      167 GAACATCACTTGTACTTGTACTTACTACAGAAATCCGGAAGAGTAAGACATGACAGAA 226
Qy      137 ArgGlyTyrSerHisSerCysLysLeuGlnHisCysAspPhePheLeuSerTyr 156
Db      227 CCGGTGTACTTCACTGCGCTGTGTGACACAGAGATTTGACATGTTCAATGACCAAGTAC 286
Qy      157 SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 176
Db      287 AGCAATCAAGAGAACCAACACTGTGTGCTTACCGGCTGACAGACATGAAAGATGAT 346
Qy      177 ProThrCysLysThrLysGluPheTyrPalaThrIleLeuAspSerAlaGlyProLeuPro 196
Db      347 GCAAGCTCAAAAGACAAAGAAATTTGGCTGACAAATTTGGATTCAACAGGCGCTCTTCT 406
Qy      197 AspTyrThrProProGluLysPheSerPheGluSerThrThrGlyPheThrLysTyrGly 216

```

Db 407 GATTACATTCCCGAGAGTGTTCCTTTAGAGAGCTCCAGGGGTTTACTATAGG 466
 Qy 217 MetLeuTyLysProHisAspLeuGlnProGlyLysLysTyLysProThrValLeuPheIle 236
 Db 467 ATGATGTAACAACCTCACAATCTGCACCTGGAAAGAGTACCTTACTGTATCTTCAATC 526
 Qy 237 TTTGlyGlyLysProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyLysPheArg 256
 Db 527 TATGAGAGGCGCTCAGGTGAGCTAGTGAACAATCAATTTAAAGAGACTAAATATTTCCCA 586
 Qy 257 LeuAsnThrLeuAlaSerLeuGlyTyLysValValValIleAspAsnArgLysSerCys 276
 Db 587 TTGAACACCTTGCGCTCTTAAAGCTACCTGTGCTGTATTTAGACAACCGGGGCTCTGC 646
 Qy 277 HisArgGlyLeuLysPheGlnGlyValAphelysTyLysMetGlyGlnIleGlnIleAsp 296
 Db 647 GACCGAGGCGCTGAGTTTGAAGAGGCTTTAAATACAAATAGGACAATATGAAATTTGAT 706
 Qy 297 AspGlnValGlu-GlyLeuGlnTyLysLeuAlaSerArgTyLysAspPheIleAspLeuAspArg 316
 Db 707 GACCAAGTGAAGGCTGACACTACTTGGCGCTCCAGTATGACTTCAATGATTTGGCTTGG 766
 Qy 316 GValGlyIleHisGlyTyLysSerTyLysGlyTyLys--LeuSerLeuMetAlaLeuMetGly 335
 Db 767 GGTGGCATTCATGGGTGCTCTTAAAGAGGCGTAACTCTCTTATAGGCGTTTAAATGCC 826
 Qy 335 nArgSer AspIlePheArgValAlaIleAlaGlyAla-ProValThrLeuTrp--Ile 353
 Db 827 GAGGTGAGATATCTTCAGAGGTGGCCATCGCGATCACCGGTAAOCTGAGGAGAGATT 886
 Qy 354 PheTyLysPheThrGly 358
 Db 887 CCTATGACACGAGGT 901
 RESULT 13 674 bp mRNA linear EST 20-OCT-2000
 BE888665 LOCUS 601513061FI NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3914582 5',
 DEFINITION mRNA sequence.
 ACCESSION BE888665
 VERSION BE888665.1 GI:10345197
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NIH-MGC http://mgi.mgi.nhl.gov/.
 AUTHORS NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapds-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LIML at:
 http://image.llnl.gov
 Plate: LIM9736 row: 1 column: 15
 High quality sequence stop: 662.
 Location/Qualifiers
 1..674
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3914582"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_idb="NIH_MGC_71"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.1 Kb."

BASE COUNT 185 a 143 c 148 g 198 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5,52e-118 Length: 674
 Score: 1117.00 Matches: 219
 Percent Similarity: 96.90% Conservative: 0
 Best Local Similarity: 96.90% Mismatches: 6
 Query Match: 57.79% Indels: 4
 DB: 10 Gaps: 0

US-10-070-464-7 (1-360) x BE888665 (1-674)

Qy 115 ProLeuGlnHisIleLeuTyValValSerTyValAsnProGlyValThrArgLeu 134
 Db 2 CCTTTAGAGCATCAGCTGACCTGACCTAGTACGTTAACTCCGAGAGGTGACAGGCTG 61
 Qy 135 ThrAspArgGlyTyLysSerHisSerCysIleSerGlnHisCysAspPheIleSer 154
 Db 62 ACTGACCGTGGCTACTACATCTTCTGTCACATGACGACACTGTGACTTCTTATAGT 121
 Qy 155 LysTySerAsnGlnLysAsnProHisCysValSerLeuTyLysLeuSerSerProGln 174
 Db 122 AAGTATAGTAACCAAGAGATCCACACTGTGATCCCTTACAGACTATCAAGCTCGAA 181
 Qy 175 AspAspProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyPro 194
 Db 182 GATGACCCCAACTGGCAAAACAAAGAAATTTGGGCCACATTTTGATTCAGAGGTCTCT 241
 Qy 195 LeuProAspTyLysThrProGlnIlePheSerPheGluSerThrThrGlyPheThrLeu 214
 Db 242 CTTCCTGACTATACCTCCCAAAATTTCTCTTTGAAAGTACTGATTTACATTG 301
 Qy 215 TyGlyMetLeuTyLysProHisAspLeuGlnProGlyLysLysTyLysProThrValLeu 234
 Db 302 TATGGATGCTCTACAGAGCTCATGATCTACAGCTCGAAAGAAATATCTTACTGTGCTG 361
 Qy 235 PheIleTyLysGlyLysProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyLys 254
 Db 362 TTCATATATGATGCTCTCAGGTGCAAGTGTGAATATCGGTTTAAAGAGTCAAGTAT 421
 Qy 255 PheArgLeuAsnThrLeuAlaSerLeuGlyTyLysValValValIleAspAsnArgGly 274
 Db 422 TTCGGTTGAATACCTTACCTCTCTAGGTTATGTGTGTATGATGATACACAGGGGA 481
 Qy 275 SerCysHisArgGlyLeuLysPheGlnGlyValAphelysTyLysMetGlyGlnIleGlu 294
 Db 482 TCCTGTACACGAGGCGCTTAAATTTGAAGGCGCTTTAAATATGAATGAGTCAACATAGAA 541
 Qy 295 IleAspAspGlnValGlnGly-LeuGlnTyLysLeuAlaSerArgTyLysAspPheIleAsp 314
 Db 542 ATTTGACGATCAGGTGAGAGAACTCCAAATATCTAGCTTCCGATATGATTTCTTACTT 601
 Qy 314 uAspArgValGlyIleHisGlyTyLysSerTyLysGlyTyLysLeuSerLeuMetAlaLeu 334
 Db 602 AGATGTGTGGGATCCACAGG-TGGTCTATGAGAGATCTC-TCCCTATGGCAATTAAT 659
 Qy 334 cGlnArgSerAspIle 339
 Db 660 GCAAG-TCAGATATC 674
 RESULT 14 1090 bp mRNA linear EST 12-MAR-2002
 BM910838 LOCUS 6615934 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454402
 DEFINITION AGENCOURT 6615934 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454402
 5', mRNA sequence.
 ACCESSION BM910838
 VERSION BM910838.1 GI:19361217
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1090)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L10M1949 row: h column: 19
 High quality sequence stop: 622.
 Location/Qualifiers

FEATURES

source

1..1090
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5454402"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC_98"
 /note="Organ: brain; Vector: pOTB7, Site_1: XhoI, Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: This is a NIH_MGC Library."

BASE COUNT 317 a 254 c 240 g 278 t 1 others

ALIGNMENT Scores:

Pidm. No.: 1.11e-116 Length: 1090
 Score: 1108.50 Matches: 237
 Percent Similarity: 77.92% Conservative: 10
 Best Local Similarity: 74.76% Mismatches: 54
 Query Match: 57.35% Gaps: 16
 DB: 12 Gaps: 4

US-10-070-464-7 (1-360) x BM910838 (1-1090)

QY 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGluGluPheAspArg 20
 DB 82 GAAGAGATGCGAGATCGAGTGCCTTGTCTCCAGAGAAATTGATGCA 141
 QY 21 TysSerGlyIYrTPTTPCySProlYsAlaGluThrThrProSerGlyGlyLysIleLeu 40
 DB 142 TATTCTGGCTATGTGTGTGTCCTCAAAAGCTGAAACACTCCACAGTGTGTAAATTCCTT 201
 QY 41 ArgIleLeuTYrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
 DB 202 AGAATTCATATGAAAGAAATGATGATCTGAGGTGGAAATTTATTCAGTTACATCCCT 261
 QY 61 MetLeuGluThrArgAlaAspSerPheArgTYrProLYsThrGlyThrAlaAsnPro 80
 DB 262 ATGTTGGAACAAGAGAGGACAGATTCAATCCGTTATCTTAAACGAGGATTTCAAGTGT 321
 QY 81 LysValThrPheLysMetSer-----GluIleMetIleAspAlaGlu 94
 DB 322 CCTATCAAGAAGAGAGATGCAATTACAGAGTGTGATGAGATGCTTGTGGCCGCGCATGCA 381
 QY 95 GlyArgIleIleValAspGluValArgArgLeuValTYrPheGluGlyThrLysAspSer 114
 DB 382 TCTATATATCCAGTTGATGAGATCGAAGGCTGGTATTTTGAAGGCAACAAAGATCTCC 441
 QY 115 ProLeuGluHisHisLeuTYrValValSerTYrValAsnProGlyGluValThrArgLeu 134
 DB 442 CCTTAGAGCATCACTGTACGTAGTCAATTACGTAATCTCGAGAGAGGTGACAAAGCTGTG 501

QY 135 ThrAspArgGlyTYrSerHisSerCysCysIleSerGlnHisCysAspPheIleSer 154
 DB 502 ACTGACCGTGGCTACTACATCTTGTGTCATCACTGACAGCACTGACATCTTTATAGT 561
 QY 155 LysTYrSerAmGlnLysAsnProHisCysValSerLeuTYrLysLeuSerSer-ProG1 174
 DB 562 AAGTATGATTAACCAAGAAATTCACACGTGTGTCCTTTAAAGCTATCAAGTCCCTGA 621
 QY 174 uAspAspProThrCysLysThrLysGluPheThrAlaThrIleLeuAspSerAlaGly-P 194
 DB 622 AGATATCCCACTTCGCAAAACAAAGAAATTTGGGCCACCATTTTGATTCAGCAGGNC 681
 QY 194 ToleuProAspTYrThrProProGluIle-PheSerPheGluSerThrGly-PheTh 213
 DB 682 CTCTCTGACTATCTCTCCTCCAGAAATTTTCTTTTGAAGATCACTACCGAAATTAC 741
 QY 213 rLeuTYrGlyMetLeu-TyrLysProHisAspLeuGluProGlyLysLysTYrProThrv 233
 DB 742 ATTGATAGGAGTCTCTACCAAGCTCAATGATCTTACGCTGGAAGAAATATCTACTAG 801
 QY 233 allLeuPhe-IleTYrGlyGlyProGluValGlnLeuVal-AsnAsnArgPheLysGlyVa 252
 DB 802 TGCTTGTCCATATAGGGGGGTCTCTCCGGGCGCAGTTGGCGGAAAAACCGCTTAAAGAT 861
 QY 252 lLysTYrPheArgLeuAsnThrLeuAlaSerLeu---GlyTYrValValValIleAs 271
 DB 862 CAAGTATTTCCCTCTGAAAACCTTCMAACCTCTTAAAGTATTGGGGTGGCAGCATAC 921
 QY 271 PAsnArgIly---SerCysHisArgGly---LeuLysPheGluGluValaPheLysTYr 289
 DB 922 ACMAAGCGCGGTCTGTCCCAACCCAGCCGTTATACATTTTGAAGGGGCTTTAAATTA 981
 QY 289 smetGlyGlnIleGluIleAspAspGluValaGluGly 301
 DB 982 CAAAGCGCCGCAACGACACAGACTACTCGCGGGT 1018

RESULT 15

BG165205 639 bp mRNA linear EST 06-FEB-2001
 LOCUS 602344074r1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4453987 5',
 DEFINITION mRNA sequence.

ACCESSION BG165205.1 GI:12671908

VERSION BG165205.1 GI:12671908

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE 1 (bases 1 to 639)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L10M10245 row: d column: 20

High quality sequence stop: 639.

Location/Qualifiers

1..639

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4453987"

/tissue_type="hypernephroma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_89"

/note="Organ: Kidney; Vector: pCMV-Sport6; Site_1: NotI;

Site 2: Salt: Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."

BASE COUNT 169 a 137 c 142 g 191 t
ORIGIN

Alignment Scores:

Pred. No.:	6,03e-116	Length:	639
Score:	1099.00	Matches:	210
Percent Similarity:	98.59%	Conservative:	0
Best Local Similarity:	98.59%	Mismatches:	2
Query Match:	56.85%	Indels:	2
DB:	10	Gaps:	0

US-10-070-464-7 (1-360) x BG165205 (1-639)

```
QY      140 SerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyrSerAengIn 159
      2  TCACATTCTTGCTGCATCAGTCAGCAGCTGGACTTCTTATAGTATAGTAAACAG 61
QY      160 LysAsnProHisCysValSerLeuTyrIleLeuSerSerProGluAspAspProThrCys 179
      62 AAGAATCCACACTGTGTGCCCTTTACAAGCTATCAAGTCTGAGATGACCAACTTGC 121
QY      180 LysThrIleGluPheThrPalaThrIleLeuAspSerAlaGlyProLeuProAspTyrThr 199
      122 AAAACAAGGAATTTGGGCACACCAT-TTGGAATTCAGCAGGTCCTTCTTGCACTACT 180
QY      200 ProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyr 219
      181 CCTCCAGAAATTTTCTCTTTGAAGTACTACTGATTTACATTGATGGGATGCTTAC 240
QY      220 LysProHisAspLeuGlnProGlyLysIleTyrProThrValLeuPheIleTyrGlyGly 239
      241 AAGCCTCATGATCTACAGCCTGGAAGAAATATCTACTGTGTGTCTATATATGTGTGT 300
QY      240 ProGlnValGlnLeuValAsnAsnArgPheLysGlyValIleTyrPheArgLeuAsnThr 259
      301 CTTCAAGTGCAGTGTGTGAATATCGTGTAAAGAGTCAAGTATTTCCGCTTGAATACC 360
QY      260 LeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCysHisArgGly 279
      361 CTAGCCTCTCTAGGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
QY      280 LeuLysPheGluGlyAlaPheLysTyrIleLysMetGlyGlnIleGluIleAspAspGlnVal 299
      421 CTTAAATTTGAAGGCCCTTTAAATATTAATGGGTCAAATAGAAATTAACGATCAGGTG 480
QY      300 GluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 319
      481 GAAGGACTCCATATCTAGCTTCTGATATGATTTCACTGACTTAGATGCTGTGGGCAATC 540
QY      320 HisGlyTyrSerTyrGlyGlyTyrIleLeuSerLeuMetAlaLeuMetGlnArgSerAspIle 339
      542 CACGGCTGTCTCATGAGGATACCTCTCCGTGATGCAATATGACAGAGTCAAGATATC 600
QY      340 Phe-ArgValAlaIleAlaGlyAlaProValThrLeu 351
      601 TTCAGCGGTGCTATTGCTGGGGCCCCAGTCACTCTG 637
DB
```

Search completed: October 16, 2003, 03:25:54
Job time : 2181.86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 15:44:49 ; Search time 53.7858 Seconds
(without alignments)
2602.858 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700

Sequence: 1 MAAMETEQGVGFETADC.....HLHYIQENIGSRIALAKVI 882

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_190jun03:*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4700	100.0	882	22	AA847187	Human DPP8. Homo
2	4700	100.0	882	23	AA824170	Human DPP8. Homo
3	4700	100.0	882	23	ABG61591	Human DPP8. Homo
4	4700	100.0	882	23	AAU74749	Human DPP8. Homo
5	4700	100.0	882	23	AAU74749	Human DPP8. Homo
6	4700	100.0	882	24	ABU07720	Human DPP8. Homo
7	4700	100.0	882	24	ABU07720	Human DPP8. Homo
8	3771	80.2	724	23	AB897362	Novel human protei
9	3504	74.6	658	23	ABG61600	Human DPP8. Homo

10	3504	74.6	661	23	ABG61596	Human DPP8. Homo
11	3504	74.6	660	23	ABG61594	Human DPP8. Homo
12	3333.5	70.9	632	22	AA893565	Human DPP8. Homo
13	3235	68.9	613	23	ABG61601	Human DPP8. Homo
14	2870	61.1	863	23	ABG61592	Human DPP8. Homo
15	2870	61.1	862	23	ABG61593	Human DPP8. Homo
16	2870	61.1	862	23	ABG61602	Human DPP8. Homo
17	2870	61.1	862	23	ABG61604	Human DPP8. Homo
18	2863	60.9	862	23	AA824168	Human DPP8. Homo
19	2835	60.3	860	23	AA824171	Human DPP8. Homo
20	2833	60.3	860	23	AA824169	Human DPP8. Homo
21	2820.5	60.0	879	23	ABG61607	Human DPP8. Homo
22	2820.5	60.0	879	23	ABG61608	Human DPP8. Homo
23	2713.5	57.7	847	23	AA823875	Human DPP8. Homo
24	2547.5	54.2	580	23	AA814337	Human DPP8. Homo
25	2422	51.5	465	22	AA847189	Human DPP8. Homo
26	2413.5	51.4	757	22	AA847189	Human DPP8. Homo
27	2406	51.2	832	23	ABG61605	Human DPP8. Homo
28	2406	51.2	832	23	ABG61606	Human DPP8. Homo
29	2356.5	50.1	819	23	ABG61609	Human DPP8. Homo
30	2356.5	50.1	819	23	ABG61610	Human DPP8. Homo
31	2233.5	47.5	683	22	AA840510	Human DPP8. Homo
32	1914.5	40.7	720	21	AA841626	Human DPP8. Homo
33	1836.5	39.1	360	22	AA847190	Human DPP8. Homo
34	1808	38.5	358	23	ABG61597	Human DPP8. Homo
35	1754.5	37.3	497	23	ABG64845	Human DPP8. Homo
36	1754.5	37.3	497	23	AAU96169	Human DPP8. Homo
37	1750.5	37.2	497	23	ABG64844	Human DPP8. Homo
38	1750.5	37.2	497	23	AAU96192	Human DPP8. Homo
39	1645.5	35.0	310	22	AA847188	Human DPP8. Homo
40	1645.5	35.0	310	23	ABG61594	Human DPP8. Homo
41	1643.5	35.0	508	21	AA842928	Human DPP8. Homo
42	1599.5	34.0	1042	22	ABG60137	Human DPP8. Homo
43	1599.5	34.0	1192	22	ABG62029	Human DPP8. Homo
44	1400	29.8	518	21	AAU90239	Human DPP8. Homo
45	1391	29.6	518	23	ABG61603	Human DPP8. Homo

ALIGNMENTS

RESULT 1	
AA847187	
ID	AA847187 standard; Protein; 882 AA.
XX	
AC	AA847187;
XX	
DT	29-JUN-2001 (first entry)
XX	
DE	Human DPP8.
XX	
KM	Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
KM	dipeptidyl peptidase; DPP8; T cell; cleavage; diarrhoea;
KM	growth hormone deficiency; glucose level; mucosal regeneration;
KM	non-insulin dependent diabetes mellitus; glucose intolerance;
KM	immunopressure.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Active-site
FT	Active-site
FT	Active-site
FT	Active-site
XX	
XX	WO200119866-A1.
XX	
XX	22-MAR-2001.
XX	
PF	11-SEP-2000; 2000MO-AU01085.
XX	

```

PR 10-SEP-1999; 99AU-0002762.
PR 18-FEB-2000; 2000AU-0005709.
XX
PA (UNSY ) UNIV SYDNEY.
XX
PI Abbott CA, Gorell MD;
XX
DR WPI; 2001-281520/29.
DR N-PSDB; AAC85694.
XX
PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT substrates, identifying inhibitors of DPP8 catalytic activity which
PT have therapeutic uses, and for detecting activated T cells
XX
PS Claim 1; Fig 2; 78pp; English.
XX
CC This sequence represents human dipeptidyl aminopeptidase (DPP8).
CC DPP8 has substrate specificity for H-Ala-Pro-pNa, H-Gly-Pro-pNa and
CC H-Arg-Pro-pNa. Therefore, it is a prolyl oligopeptidase and a
CC dipeptidyl peptidase, because it is capable of hydrolysing the
CC peptide bond C-terminal to Pro in each of these compounds. DPP8
CC is homologous with human DPPIV. DPP8 is useful for cleaving a
CC substrate, and for detecting an activated T cell which involves
CC measuring the level of DPP8 gene expression in a T cell. The level
CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
CC in the cell. It is also useful for identifying a molecule capable
CC of inhibiting the cleavage of the substrate by DPP8. Molecules
CC identified as inhibiting DPP8 catalytic activity may be useful for
CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
CC in non-insulin dependent diabetes mellitus and other disorders
CC involving glucose intolerance, enhancing mucosal regeneration and
CC as immunosuppressants.
XX
SQ Sequence 882 AA;
XX
Query Match 100.0%; Score 4700; DB 22; Length 882;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MAAMETEQGVIEFETADCEENIESQDRPKLEPPYVRYVMSQKLLADTRKHYGM 60
Db 1 MAAMETEQGVIEFETADCEENIESQDRPKLEPPYVRYVMSQKLLADTRKHYGM 60
QY 61 AKAPHEMFVKRNDPDPGHSRIYLLAMSGENRENTLFYSSIPKTINRAAVLMSKPL 120
Db 61 AKAPHEMFVKRNDPDPGHSRIYLLAMSGENRENTLFYSSIPKTINRAAVLMSKPL 120
QY 121 DLFOATLDYGVMSPREELLRKRKRGIVGASVDYHQSGLFPLQAGSGIYHVVDGPG 180
Db 121 DLFOATLDYGVMSPREELLRKRKRGIVGASVDYHQSGLFPLQAGSGIYHVVDGPG 180
QY 181 FTQOPLRENLVETSCPNIRMDPKLCPADPWIAFIHSDNIMISNIVTREERLLTYVNE 240
Db 181 FTQOPLRENLVETSCPNIRMDPKLCPADPWIAFIHSDNIMISNIVTREERLLTYVNE 240
QY 241 ANMEDDARSAGVATFVLOEEDRYSGYWMCRAETTPSGKILRLIYEENDESEVEI 300
Db 241 ANMEDDARSAGVATFVLOEEDRYSGYWMCRAETTPSGKILRLIYEENDESEVEI 300
QY 301 TSPMLERRADSPRYPKGTANPKYTFPMSEIMIDABRIIDVIDKELIOPFEIFFEG 360
Db 301 TSPMLERRADSPRYPKGTANPKYTFPMSEIMIDABRIIDVIDKELIOPFEIFFEG 360
QY 361 YIARAGWTPEGKYAMSILLDRSQTRLQIVLISPFLIPEVDVMERQRLISVDPSTPL 420
Db 361 YIARAGWTPEGKYAMSILLDRSQTRLQIVLISPFLIPEVDVMERQRLISVDPSTPL 420
QY 421 IIVETTDIWINIHIDIEHVPOSHRETEFIASCKCKGFHLKYTISILSKSKRSSG 480
Db 421 IIVETTDIWINIHIDIEHVPOSHRETEFIASCKCKGFHLKYTISILSKSKRSSG 480
QY 481 GLPAPDFKCPIKEIEAITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVS 540
Db 481 GLPAPDFKCPIKEIEAITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVS 540

```

```

Db 481 GLPAPDFKCPIKEIEAITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVS 540
QY 541 YVNPGEVTRILDRGYSHSCCISQCHDFIISKYSNOKAPHCVSLYKLSPEDDPYCKTKEF 600
Db 541 YVNPGEVTRILDRGYSHSCCISQCHDFIISKYSNOKAPHCVSLYKLSPEDDPYCKTKEF 600
QY 601 WATILDSAGPLPDYTPPELIFSFEETGTFTYGLMYKPHDIQPKKPYTVLFYGGPOVL 660
Db 601 WATILDSAGPLPDYTPPELIFSFEETGTFTYGLMYKPHDIQPKKPYTVLFYGGPOVL 660
QY 661 VNNRFKGVKFRNLTLASLGVVVVINDRRSGCHGKKEGAFKYMGOIEIDDOVEGQY 720
Db 661 VNNRFKGVKFRNLTLASLGVVVVINDRRSGCHGKKEGAFKYMGOIEIDDOVEGQY 720
QY 721 LASRYDFLDRVGIHMSYGYSLMALMQRSDIFRYALAGAPVTLMIFYDTGYTERYM 780
Db 721 LASRYDFLDRVGIHMSYGYSLMALMQRSDIFRYALAGAPVTLMIFYDTGYTERYM 780
QY 781 GHPDQNEQGYLGSVAMQAEKFPSEPNRLILHGFLEENVFAPHTSILSLVRAKPYD 840
Db 781 GHPDQNEQGYLGSVAMQAEKFPSEPNRLILHGFLEENVFAPHTSILSLVRAKPYD 840
QY 841 LQIYPOERHSIRVPSGHEVHLHLHYLOENLGSRIALAKVI 882
Db 841 LQIYPOERHSIRVPSGHEVHLHLHYLOENLGSRIALAKVI 882
XX
RESULT 2
AAE24170
ID AAE24170 standard; Protein: 882 AA.
XX
AC AAE24170;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human dipeptidyl peptidase 8 (DPP8) protein.
XX
KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; Human immuno deficiency virus; HIV infection; cytostatic;
KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW antiviral; enzyme.
XX
OS Homo sapiens.
XX
PN WO200234900-A1.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-AU01388.
XX
PR 27-OCT-2000; 2000AU-0001078.
XX
PA (UNSY ) UNIV SYDNEY.
XX
PI Abbott CA, Gorell MD;
XX
DR WPI; 2002-454646/48.
DR N-PSDB; AAD38956.
XX
PT New dipeptidyl peptidase (DPP) peptides, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection.
XX
PS Example; Fig 1; 91pp; English.
XX
CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC polynucleotides encoding such proteins. The DPP peptides are useful for
CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC rejection and HIV (human immuno deficiency virus) infection. The present
CC sequence is human DPP8 protein.

```

Sequence	882 AA;
Query Match	100.0%; Score 4700; DB 23; Length 882;
Best Local Similarity	100.0%; Pred. No. 0;
Matches	882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MAAAMEEQLGVEIFETADCEENIESODRKLPEFYERYSWSQKLLADTRKHYGYM 60
DB	1 MAAAMEEQLGVEIFETADCEENIESODRKLPEFYERYSWSQKLLADTRKHYGYM 60
QY	61 AKAPHEMFVKRNDPDGPHSDRIYYLAMSSENNENTLFYSEIPKTIINRAAVLMSKPELL 120
DB	61 AKAPHEMFVKRNDPDGPHSDRIYYLAMSSENNENTLFYSEIPKTIINRAAVLMSKPELL 120
QY	121 DLFOATLDYGYMSREBELRERKRIGTVGASVDYHOGSGTFLFOAGSGIYHVKGDPGQ 180
DB	121 DLFOATLDYGYMSREBELRERKRIGTVGASVDYHOGSGTFLFOAGSGIYHVKGDPGQ 180
QY	181 FTQOPLRPLVETSCPNIRMDPKLCPADPDWIAFIHNSNDIWNISNITVREERLTYVHNE 240
DB	181 FTQOPLRPLVETSCPNIRMDPKLCPADPDWIAFIHNSNDIWNISNITVREERLTYVHNE 240
QY	241 ANNEEDARSAGVATFVLOEEFDRYSGYMWCPKAETTPSGGKILRIIYEENDESEVELIHY 300
DB	241 ANNEEDARSAGVATFVLOEEFDRYSGYMWCPKAETTPSGGKILRIIYEENDESEVELIHY 300
QY	301 TSPMLETRRADSRFRYKGTGTANPKVTFKNSEIMIDAGRIIDVIDKELIOPFEILFEGVE 360
DB	301 TSPMLETRRADSRFRYKGTGTANPKVTFKNSEIMIDAGRIIDVIDKELIOPFEILFEGVE 360
QY	361 YIARAWTPEGKAWMSIILDRSOTRIQIVLISELPIVAVDDMBEORLIESVPDVTPL 420
DB	361 YIARAWTPEGKAWMSIILDRSOTRIQIVLISELPIVAVDDMBEORLIESVPDVTPL 420
QY	421 IYEETTDIMINIDHFHVPOSGHEEIEPIFASECKTGFRHLKYITSLIKESKYRSSG 480
DB	421 IYEETTDIMINIDHFHVPOSGHEEIEPIFASECKTGFRHLKYITSLIKESKYRSSG 480
QY	481 GLPAPDFKCPKIEEIAITSGEWEVLGRHGSNIQVDEVRLLVFEESTKOSPLEHLLVYVS 540
DB	481 GLPAPDFKCPKIEEIAITSGEWEVLGRHGSNIQVDEVRLLVFEESTKOSPLEHLLVYVS 540
QY	541 YVNRPGVETLDRGVSHSCCISQHCDFISKYSNQNPCVSLYKLSSEDDPTCTKXEF 600
DB	541 YVNRPGVETLDRGVSHSCCISQHCDFISKYSNQNPCVSLYKLSSEDDPTCTKXEF 600
QY	601 MATIIDSAGPLPDYTPPEIFSEFESTTGFTLYGMLYKPHDLOPKKYPVLFTYGGPQVQL 660
DB	601 MATIIDSAGPLPDYTPPEIFSEFESTTGFTLYGMLYKPHDLOPKKYPVLFTYGGPQVQL 660
QY	661 VNNRFKGVKYYFRLNTLASLGYVVVVINDNGSCHRGKLFEGAFKRYXMGQIIEIDQVAGLQY 720
DB	661 VNNRFKGVKYYFRLNTLASLGYVVVVINDNGSCHRGKLFEGAFKRYXMGQIIEIDQVAGLQY 720
QY	721 LASRYDFILDRGIGHGSGYGYLSMALMORSIDIRVVAIAGAPVLTMTFYDGYGERM 780
DB	721 LASRYDFILDRGIGHGSGYGYLSMALMORSIDIRVVAIAGAPVLTMTFYDGYGERM 780
QY	781 GHDPDNEGQYLYGSVAMQAEKFPSEBNRLLLHGFLENVHFAHTSILSLFVRAKPYD 840
DB	781 GHDPDNEGQYLYGSVAMQAEKFPSEBNRLLLHGFLENVHFAHTSILSLFVRAKPYD 840
QY	841 LQIYPOERHSIRVPSGGEYELHLHYLOENTGSRIALAKVI 882
DB	841 LQIYPOERHSIRVPSGGEYELHLHYLOENTGSRIALAKVI 882

RESULT 3
ABG61591

ID ABG61591 standard; Protein; 882 AA.
XX
AC
XX ABG61591;

DT	12-ANG-2002 (first entry)
XX	Human DPPIV related serine protease DPP-1.
DE	Human, serine protease; dipeptidyl peptidase IV-related protein; DPP;
XX	DPP-1; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW	heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW	dyskinesia; reproductive disorder; inflammatory disorder;
KW	metabolic disorder.
OS	Homo sapiens.
PN	MO200231134-A2.
XX	18-ARR-2002.
PD	12-OCT-2001; 2001MO-US31874.
PF	12-OCT-2001; 2000US-240117P.
PR	(PERR) FERRING BV.
XX	Qi S, Akinsanya KO, Riviere PJ, Junien J;
PA	WPI; 2002-444178/47.
PI	N-PSDB; ABR33322.
DR	New dipeptidyl ⁴ peptidase IV-related proteins and nucleic acids encoding
PT	the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT	viral infections, cancers, allergies, neurological disorders, or pain
XX	Claim 17; Fig 1; 113bp; English.
XX	The present invention relates to the isolation of novel human serine
CC	proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC	proteins (DPP). The dipeptidyl peptidase IV-related proteins (DPP)
CC	and nucleic acids encoding them are useful for treating infections
CC	such as fungal, bacterial, protozoan and viral infections, particularly
CC	infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC	pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC	bulimia, Parkinson's disease, acute heart failure, hypertension,
CC	hypertension, urinary retention, osteoporosis, angina pectoris,
CC	stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC	psychotic and neurological disorders (e.g. anxiety, dementia, or
CC	schizophrenia), and dyskinesias. These may also be used in discovering
CC	therapeutic agents for the treatment of reproductive, inflammatory and
CC	metabolic disorders. ABG61591-ABG61612 represent human DPP proteins.
XX	Sequence 882 AA;
SO	Query Match
XX	Best Local Similarity 100.0%; Score 4700; DB 23; Length 882;
XX	Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MAAAMEEQLGVEIFETADCEENIESODRKLPEFYERYSWSQKLLADTRKHYGYM 60
DB	1 MAAAMEEQLGVEIFETADCEENIESODRKLPEFYERYSWSQKLLADTRKHYGYM 60
QY	61 AKAPHEMFVKRNDPDGPHSDRIYYLAMSSENNENTLFYSEIPKTIINRAAVLMSKPELL 120
DB	61 AKAPHEMFVKRNDPDGPHSDRIYYLAMSSENNENTLFYSEIPKTIINRAAVLMSKPELL 120
QY	121 DLFOATLDYGYMSREBELRERKRIGTVGASVDYHOGSGTFLFOAGSGIYHVKGDPGQ 180
DB	121 DLFOATLDYGYMSREBELRERKRIGTVGASVDYHOGSGTFLFOAGSGIYHVKGDPGQ 180
QY	181 FTQOPLRPLVETSCPNIRMDPKLCPADPDWIAFIHNSNDIWNISNITVREERLTYVHNE 240
DB	181 FTQOPLRPLVETSCPNIRMDPKLCPADPDWIAFIHNSNDIWNISNITVREERLTYVHNE 240

Good for
diagnosis


```
QY 241 ANNEEDARSAGVATFVLQEEFDRIQYVWCPKAEPTSGKILRLIYEENDESEVEIHHV 300
DB 241 ANNEEDARSAGVATFVLQEEFDRIQYVWCPKAEPTSGKILRLIYEENDESEVEIHHV 300
QY 301 TSPMLFTRRADSFYRPTGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFILFEGVE 360
DB 301 TSPMLFTRRADSFYRPTGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFILFEGVE 360
QY 361 YIARAGTPEGKYVMSILDRSOTRLOIVLISPELFIPEDDVNERORLISVDSVTP 420
DB 361 YIARAGTPEGKYVMSILDRSOTRLOIVLISPELFIPEDDVNERORLISVDSVTP 420
QY 421 IYEETTDIWINIHDFHVPQSHHEEIEFIASECTGFRHLKYTISILKESYKXSSG 480
DB 421 IYEETTDIWINIHDFHVPQSHHEEIEFIASECTGFRHLKYTISILKESYKXSSG 480
QY 481 GLPAPSPFKCPIKEEIAITSGEMVLRHGSNIQVDYRRLVYREGTKDSLEHHLVVS 540
DB 481 GLPAPSPFKCPIKEEIAITSGEMVLRHGSNIQVDYRRLVYREGTKDSLEHHLVVS 540
QY 541 YVNGEVTRLTDRGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTKEF 600
DB 541 YVNGEVTRLTDRGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTKEF 600
QY 601 WATILDSAGPLPDYTPPEIIFSGESTTGFTLYGMLYKPHDIQPKKYPVTLFIYGGPQVL 660
DB 601 WATILDSAGPLPDYTPPEIIFSGESTTGFTLYGMLYKPHDIQPKKYPVTLFIYGGPQVL 660
QY 661 VNNRFKVKYFRLNTLASLGVVVVINDRSGCHRGKPEGAFFKMKQIETDDOVELQY 720
DB 661 VNNRFKVKYFRLNTLASLGVVVVINDRSGCHRGKPEGAFFKMKQIETDDOVELQY 720
QY 721 LASRYDEIDLDVRIHMSYGYLSLWALMORSIDIFFVALAGAVTLIMIFDTYTERYM 780
DB 721 LASRYDEIDLDVRIHMSYGYLSLWALMORSIDIFFVALAGAVTLIMIFDTYTERYM 780
QY 781 GHPDQNEQGYLLGSVAMQAEKFPSEPNRLILHGFLEDBNVHFAHTSILSLFVAGKPYD 840
DB 781 GHPDQNEQGYLLGSVAMQAEKFPSEPNRLILHGFLEDBNVHFAHTSILSLFVAGKPYD 840
QY 841 LQIYPOERHSIRVPSGEHYELHLHYLOENLGSRIALKVI 882
DB 841 LQIYPOERHSIRVPSGEHYELHLHYLOENLGSRIALKVI 882
```

RESULT 4
AAU74749
ID AAU74749 standard; Protein: 882 AA.

AAU74749;
09-APR-2002 (first entry)
DE Human protease PRTS-9 protein sequence.

Human, protease; PRTS; gastrointestinal; Crohn's disease; cancer;
cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
cell proliferative disorder; developmental disorder; epilepsy;
Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
reproductive disorder; endometriosis.

OS Homo sapiens.
XX
XX WO200198468-A2.
XX
XX PD 27-DEC-2001.
XX
XX PR 13-JUN-2001; 2001WO-US19178.
XX
XX PR 16-JUN-2000; 2000US-212336P.
XX
XX PR 22-JUN-2000; 2000US-213955P.
XX
XX PR 29-JUN-2000; 2000US-215396P.

```
PR 07-JUL-2000; 2000US-216821P.  
PR 14-JUL-2000; 2000US-218946P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX yue H, Elliott vs, Gandhi AR, Lal P, Au-young J, Tribouley CM,  
PI Deleagane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA,  
PI Wajia NK, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT;  
PI Azimzai Y, Lu Y, Runkumar J, Xu Y, Reddy R, Das D, Kearney L,  
PI Kalltick DA;  
XX  
XX WPI; 2002-090437/12.  
DR N-PSDB; ABK12892.  
DR  
XX  
XX Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful  
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.  
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell  
PT proliferative (e.g. cancer) disorders -  
XX  
XX Claim 1; Page 140-142; 177pp; English.  
XX  
XX The present invention relates to twenty one new human proteases,  
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and  
CC polypeptides of the invention are useful in the diagnosis, treatment and  
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and  
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and  
CC myocardial infarction, autoimmune/inflammatory e.g. acquired  
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell  
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker  
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.  
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and  
CC endometriosis disorders. Numerous other examples of each disorder are  
CC given in the specification. The present protein sequence represents  
XX the human protease PRTS-9 protein of the invention.  
XX  
SQ Sequence 882 AA;  
Query Match 100.0%; Score 4700; DB 23; Length 882;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAAMETEQLGVEIFETADCEENIESODRPLEPFYERYYSWQKLKLDTRKYHYGM 60  
DB 1 MAAMETEQLGVEIFETADCEENIESODRPLEPFYERYYSWQKLKLDTRKYHYGM 60  
QY 61 AKAPHDPMFYKRPNDPOGPHSDRIYILAMSGENRENTLFYSEIPTINRAVLM.SMKPL 120  
DB 61 AKAPHDPMFYKRPNDPOGPHSDRIYILAMSGENRENTLFYSEIPTINRAVLM.SMKPL 120  
QY 121 DLFOATLDYGMYSREBELREKRIGTVGLASYDYHOGSGTFLFOAGSGIYHVADGGPQG 180  
DB 121 DLFOATLDYGMYSREBELREKRIGTVGLASYDYHOGSGTFLFOAGSGIYHVADGGPQG 180  
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDMIAFIHNSDIWISNIVTREBRRLTYVANEL 240  
DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDMIAFIHNSDIWISNIVTREBRRLTYVANEL 240  
QY 241 ANNEEDARSAGVATFVLQEEFDRIQYVWCPKAEPTSGKILRLIYEENDESEVEIHHV 300  
DB 241 ANNEEDARSAGVATFVLQEEFDRIQYVWCPKAEPTSGKILRLIYEENDESEVEIHHV 300  
QY 301 TSPMLFTRRADSFYRPTGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFILFEGVE 360  
DB 301 TSPMLFTRRADSFYRPTGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFILFEGVE 360  
QY 361 YIARAGTPEGKYVMSILDRSOTRLOIVLISPELFIPEDDVNERORLISVDSVTP 420  
DB 361 YIARAGTPEGKYVMSILDRSOTRLOIVLISPELFIPEDDVNERORLISVDSVTP 420  
QY 421 IYEETTDIWINIHDFHVPQSHHEEIEFIASECTGFRHLKYTISILKESYKXSSG 480  
DB 421 IYEETTDIWINIHDFHVPQSHHEEIEFIASECTGFRHLKYTISILKESYKXSSG 480
```

```

OY 481 GLPAPSDKCPKEEIAITSGEMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLVYVS 540
DB 491 GLPAPSDKCPKEEIAITSGEMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLVYVS 540
OY 541 YVNPGEVRLTDRGYSHSCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTKEF 600
DB 541 YVNPGEVRLTDRGYSHSCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTKEF 600
OY 601 WATILDSAGPLDPDYPPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFTYGGPOVQL 660
DB 601 WATILDSAGPLDPDYPPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFTYGGPOVQL 660
OY 661 VNNRKGKGVKFRPLNTLASIGYVVVVIDNRGSGHRLKFEKGAFKYMGGIIDDQVEGLQY 720
DB 661 VNNRKGKGVKFRPLNTLASIGYVVVVIDNRGSGHRLKFEKGAFKYMGGIIDDQVEGLQY 720
OY 721 LASRDFIDLDVGIHGSYSGYLSLMALMORSDFRVAIAGAPVTLMIFYDTGTERYM 780
DB 721 LASRDFIDLDVGIHGSYSGYLSLMALMORSDFRVAIAGAPVTLMIFYDTGTERYM 780
OY 781 GHPDNEOGYVYGSVAMQAEKFPSPBNRLLLHGFIDENVHFAHTSILLSPVRAKGPYD 840
DB 781 GHPDNEOGYVYGSVAMQAEKFPSPBNRLLLHGFIDENVHFAHTSILLSPVRAKGPYD 840
OY 841 LOIYPOERHSIRVPESGEHYEHLHYLOENIGSRILAAKVI 882
DB 841 LOIYPOERHSIRVPESGEHYEHLHYLOENIGSRILAAKVI 882

RESULT 5
AAG78415
ID AAG78415 standard; Protein; 882 AA.
AC AAG78415;
DT 12-APR-2002 (first entry)
XX
DE Amino acid sequence of 21953 human prollyl oligopeptidase.
XX
KW 21953 prollyl oligopeptidase; antibody; proline; endopeptidase;
KW cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
KW diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW Grave's disease; neuronal disorder; demyelinating disease.
XX
OS Homo sapiens.
XX
PN WO200179473-A2.
XX
PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-US40483.
XX
PR 18-APR-2000; 2000US-197508P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Meyers RA, Williamson M;
XX
DR MPI; 2002-034353/04.
DR N-PSDB; AAH99934.
XX
PT New polypeptides 21953, member of human prollyl oligopeptidase family,
XX useful as diagnostic targets and therapeutic agents for controlling
XX cancer, lymphoma and leukemia -
XX
PS Claim 1; Page 102-103; 12pp; English.
XX
CC This invention relates to an isolated 21953 human prollyl
CC oligopeptidase, which is cytosstatic, antidiabetic, antiarthritic,
CC neuroprotective, antithyroid, dermatological, antipsoriatic,
CC antiasthmatic, ophthalmological, antiinflammatory, nootropic,

```

```

CC antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
CC antianginal, cardiant, antiatherosclerotic, anorectic and
CC metabolic in its action. Uses include gene therapy, expression or
CC activity of 21953 protein modulator, it is useful for identifying a
CC compound which binds to it and can be used in preventing, treating a
CC or detecting a cellular proliferative or differentiative disorder.
CC The 21953 molecules can act as novel diagnostic targets and therapeutic
CC agents for controlling disorders associated with the aberrant activity
CC or degradation of peptide hormones e.g., disorders associated with cell
CC differentiation and proliferation such as cancer, immune function,
CC reproductive, neurological and cardiovascular function. The 21953
CC molecules are thus useful for treating and preventing cellular
CC proliferative and differentiative disorders, haematopoietic neoplastic
CC disorders, immune disorders such as autoimmune diseases, diabetes
CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC neuronal disorders, demyelinating diseases, vascular disorders and
CC metabolism or pain disorders. This sequence represents the amino
CC acid sequence of 21953 human prollyl oligopeptidase.
XX
SQ Sequence 882 AA;
Query Match 100.0%; Score 4700; DB 23; Length 882;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MAAAMETSQLGVEIFETADCEENIESODRPKLEPFYVERYSOLKKLLADTRKYGVM 60
DB 1 MAAAMETSQLGVEIFETADCEENIESODRPKLEPFYVERYSOLKKLLADTRKYGVM 60
OY 61 AKAPHDPMVRKNDPDGPHSDRIYYLAMSGERENTLFPSELPKTNRAANLMSKPLL 120
DB 61 AKAPHDPMVRKNDPDGPHSDRIYYLAMSGERENTLFPSELPKTNRAANLMSKPLL 120
OY 121 DLFOATLDYGMYSREELLREKRIGTVGIASVDYHOGSGTFLFOAGSGIYHVKDGPOG 180
DB 121 DLFOATLDYGMYSREELLREKRIGTVGIASVDYHOGSGTFLFOAGSGIYHVKDGPOG 180
OY 181 FTQGPRLRNLVETSCPNIRMDPKLCPADPDWIAFTHSNDIWSNTVTRERRLTYVHNEI 240
DB 181 FTQGPRLRNLVETSCPNIRMDPKLCPADPDWIAFTHSNDIWSNTVTRERRLTYVHNEI 240
OY 241 ANMEDDASAGYATVLOEFPRIYGYWCPAEFTTPSGKTLRLIYENDSEVEITIV 300
DB 241 ANMEDDASAGYATVLOEFPRIYGYWCPAEFTTPSGKTLRLIYENDSEVEITIV 300
OY 241 ANMEDDASAGYATVLOEFPRIYGYWCPAEFTTPSGKTLRLIYENDSEVEITIV 300
DB 241 ANMEDDASAGYATVLOEFPRIYGYWCPAEFTTPSGKTLRLIYENDSEVEITIV 300
OY 301 TSPMLETRRADSEFRPKGTANPKVTFPKMSEIMIDABGRIDVIDKELIQPEILLFEGVE 360
DB 301 TSPMLETRRADSEFRPKGTANPKVTFPKMSEIMIDABGRIDVIDKELIQPEILLFEGVE 360
OY 361 YIARAGWTPBCKYANSLIDRSQTRLQIVLISPPELLPYEDDVMERORLIESVDPSTPL 420
DB 361 YIARAGWTPBCKYANSLIDRSQTRLQIVLISPPELLPYEDDVMERORLIESVDPSTPL 420
OY 421 IYIETTDIMINIHIFHVFPQSHEEIEFTFASCKTGFRLLYKITSILKSKYKRSRG 480
DB 421 IYIETTDIMINIHIFHVFPQSHEEIEFTFASCKTGFRLLYKITSILKSKYKRSRG 480
OY 481 GLPAPSDKCPKEEIAITSGEMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLVYVS 540
DB 481 GLPAPSDKCPKEEIAITSGEMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLVYVS 540
OY 541 YVNPGEVRLTDRGYSHSCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTKEF 600
DB 541 YVNPGEVRLTDRGYSHSCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTKEF 600
OY 601 WATILDSAGPLDPDYPPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFTYGGPOVQL 660
DB 601 WATILDSAGPLDPDYPPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFTYGGPOVQL 660
OY 661 VNNRKGKGVKFRPLNTLASIGYVVVVIDNRGSGHRLKFEKGAFKYMGGIIDDQVEGLQY 720
DB 661 VNNRKGKGVKFRPLNTLASIGYVVVVIDNRGSGHRLKFEKGAFKYMGGIIDDQVEGLQY 720

```

QY 721 LASRYDFIDLRVGHGMSYGYLSLMALMORSDFRVALGAPVTLMIFDTGTERYM 780
 DB 721 LASRYDFIDLRVGHGMSYGYLSLMALMORSDFRVALGAPVTLMIFDTGTERYM 780
 QY 781 GHPDNEGGYVLSVMAQAEKFPSEPNRLLLHGFIDENVAHFAHTSILLSPVLRAGKPYD 840
 DB 781 GHPDNEGGYVLSVMAQAEKFPSEPNRLLLHGFIDENVAHFAHTSILLSPVLRAGKPYD 840
 QY 841 LQIYPOERHSIRVPSGEHVELHLMYLOENLGSRIAALKVI 882
 DB 841 LQIYPOERHSIRVPSGEHVELHLMYLOENLGSRIAALKVI 882

RESULT 6

ABU07720
 ID ABU07720 standard; Protein; 882 AA.

XX AC ABU07720;

XX DT 19-MAY-2003 (first entry)

XX DE Human serine protease HIPHM46.

XX KW Human; enzyme; HIPHM46; serine protease; gene therapy; osteoarthritis;
 KW serine protease activity modulation; dipeptidyl peptidase activity;
 KW musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
 KW Alzheimer's disease; paraspinal nuclear palsy; Huntington's disease;
 KW amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
 KW irritable bowel syndrome; type I diabetes; fecal incontinence; anaemia;
 KW haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;
 KW colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
 KW multiple sclerosis.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers

FT Region 259..260
 FT /note= "Paired glutamates of the beta propeller domain"

FT Active-site 739
 FT /label= Catalytic_serine_residue

FT Active-site 817
 FT /label= Catalytic_aspartate_residue

FT Active-site 849
 FT /label= Catalytic_hisidine_residue

XX GB2374869-A.

XX PD 30-OCT-2002.

XX PF 22-JAN-2002; 2002GB-0001404.

XX PR 23-JAN-2001; 2001GB-0001760.

XX PA (GLAXO) GLAXO GROUP LTD.

XX PI Edbrooke MR, Lewis AP;

XX WPI; 2003-150703/15.

XX DR N-PSDB; ABX12255.

PT Identifying modulators of serine protease activity useful for treating
 PT musculoskeletal diseases, by contacting cell expressing a novel serine
 PT protease polypeptide with a compound and monitoring serine protease
 PT activity -

XX Claim 10; Page 26-29; 38pp; English.

XX The invention relates to a method of identifying a substance that
 CC modulates serine protease activity, comprising contacting a cell such as
 CC a neuronal cell, lung cell, intestinal cell or a cell infected with a
 CC virus, expressing a serine protease polypeptide (HIPHM 46), or its
 CC variant having dipeptidyl peptidase activity, or a serine protease
 CC isolated from the cell with a test substance and monitoring for serine

CC protease activity. The method is useful for identifying a substance that
 CC modulates serine protease activity. A modulator of the serine protease is
 CC useful in the manufacture of a medicament for treatment or prophylaxis of
 CC a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
 CC infection, Alzheimer's disease, paraspinal nuclear palsy, myotonic
 CC dystrophy, Huntington's disease or amyotrophic lateral sclerosis.
 CC Additional disease that may be treated using modulators of the serine
 CC protease include malabsorption syndromes, irritable bowel syndrome, lung
 CC disease, type I diabetes, fecal incontinence, haemorrhoids, proctitis,
 CC rectal polyps, small bowel tumours, colorectal tumours, anaemia,
 CC dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
 CC sclerosis. The present sequence represents the amino acid sequence of the
 CC human serine protease HIPHM46.

XX SQ Sequence 882 AA;

Query Match 100.0%; Score 4700; DB 24; Length 882;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAAMETBOLGVEIFETADCEENISQDRPKLEPFVERYSQKLKLLADTRKHGYMM 60
 DB 1 MAAAMETBOLGVEIFETADCEENISQDRPKLEPFVERYSQKLKLLADTRKHGYMM 60
 QY 61 AKAHPDMFKANDPDGPHSDRIYYLAMSGENRENTLFYSEI PKTINRAAVMLSKPPL 120
 DB 61 AKAHPDMFKANDPDGPHSDRIYYLAMSGENRENTLFYSEI PKTINRAAVMLSKPPL 120
 QY 121 DLFQATLDYGMYSREBELRERKIGTVGASVDYHQSGLTFLFQAGSIIYVKDGPQG 180
 DB 121 DLFQATLDYGMYSREBELRERKIGTVGASVDYHQSGLTFLFQAGSIIYVKDGPQG 180
 QY 181 FPOQPLRPVLVETSCPNIRMDPKLCPADPDWIAFHSNDIMVSNVTRERRLTYVHNEL 240
 DB 181 FPOQPLRPVLVETSCPNIRMDPKLCPADPDWIAFHSNDIMVSNVTRERRLTYVHNEL 240
 QY 241 ANMEEDASAGATVLOEFPDYGYYWCPKAFETPSGKILRIYEENDESEVEIIHY 300
 DB 241 ANMEEDASAGATVLOEFPDYGYYWCPKAFETPSGKILRIYEENDESEVEIIHY 300
 QY 301 TSPMLETRRADSFRYPKGTANPKYTFKMSIIMDAEGRIIDVIDKELIQPEILLFEGVE 360
 DB 301 TSPMLETRRADSFRYPKGTANPKYTFKMSIIMDAEGRIIDVIDKELIQPEILLFEGVE 360
 QY 361 YIARAGMPEGKYAMSILLDRSOTRLOVLLISPELFIPEEDVMERORLIESVPSVTPL 420
 DB 361 YIARAGMPEGKYAMSILLDRSOTRLOVLLISPELFIPEEDVMERORLIESVPSVTPL 420
 QY 421 ILYEETTDIWINIHIDIFHVFQSHHEEIEFIPASECKTFRHLKYITSILKSKYKRSRG 480
 DB 421 ILYEETTDIWINIHIDIFHVFQSHHEEIEFIPASECKTFRHLKYITSILKSKYKRSRG 480
 QY 481 GLPAPSDPKCPKKEIATITSGMEVVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVVS 540
 DB 481 GLPAPSDPKCPKKEIATITSGMEVVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVVS 540
 QY 541 YNPGEVRLTRDRGSHSCCISQHCDFIISKXSNXNCPVSLYKLSPEDDPTCKTEF 600
 DB 541 YNPGEVRLTRDRGSHSCCISQHCDFIISKXSNXNCPVSLYKLSPEDDPTCKTEF 600
 QY 601 WATILDSAGPLPDYTPPEIFSFESTTGTLYGMLTKPHDLOPKKYPTVLFIYGGPOVL 660
 DB 601 WATILDSAGPLPDYTPPEIFSFESTTGTLYGMLTKPHDLOPKKYPTVLFIYGGPOVL 660
 QY 661 YNNRFGKYKFRNLNTLASLGYVVVVITNDRSGCHRLKKEGAFKYMGGIETDDQVEGLQY 720
 DB 661 YNNRFGKYKFRNLNTLASLGYVVVVITNDRSGCHRLKKEGAFKYMGGIETDDQVEGLQY 720
 QY 721 LASRYDFIDLRVGHGMSYGYLSLMALMORSDFRVALGAPVTLMIFDTGTERYM 780
 DB 721 LASRYDFIDLRVGHGMSYGYLSLMALMORSDFRVALGAPVTLMIFDTGTERYM 780
 QY 781 GHPDNEGGYVLSVMAQAEKFPSEPNRLLLHGFIDENVAHFAHTSILLSPVLRAGKPYD 840

DB 781 GHPDNEQGYLGSVAMQAEKFPSEPNRLLLHGFLENVHFAHTSILSLFLVAKGKPYD 840

QY 841 LQIYPOERHSIRVPESGEHEHLHLHYLQENLGSRIALAKVI 882

DB 841 LQIYPOERHSIRVPESGEHEHLHLHYLQENLGSRIALAKVI 882

RESULT 7
ABB97361
ID ABB97361 standard; Protein; 782 AA.
XX
AC ABB97361;
XX
DT 27-JUN-2002 (first entry)
XX
DE Novel human protein SEQ ID NO: 629.
XX

KM Human; antihaemic; vulnerary; antiinflammatory; immunomodulator;
KM antifertility; cerebroprotective; cytosatic; rheumatic; gene therapy;
KM neuroprotective; antiparkinsonian; protein therapy; EST;
KM expressed sequence tag.

XX Homo sapiens.
XX
PN M0200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US26015.
XX
PR 11-SEP-2000; 2000US-0659671.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Dermanac RT, Zhang QA, Ren F;
XX
DR WPI; 2002-292408/33.
XX
DR N-PSDB; ABB97361.
XX

PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis -
XX
XX
PS Example 2; SEQ ID NO 629; 509bp; English.
XX

CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention.
XX
SQ Sequence 782 AA.

Query Match 87.6%; Score 4118; DB 23; Length 782;
Best Local Similarity 88.7%; Pred. No. 0;
Matches 782; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 MAAAMTBQLGVIFETADCEENIESQDRPKLPPFYERYWSQKLKLADTRKYGYWM 60
DB 1 MAAAMTBQLGVIFETADCEENIESQDRPKLPPFYERYWSQKLKLADTRKYGYWM 60
QY 61 AKAPHDPMFKRDPDGPSPDRITVYLAMSGENENTLFSYIKTINRAAVLMSKPL 120
DB 61 AKAPHDPMFKRDPDGPSPDRITVYLAMSGENENTLFSYIKTINRAAVLMSKPL 120
QY 121 DLFOATLDYGMYSREBELLRERKRIQTVGIASDYHOGSGTFLFOAGSGIYHVKDGPQG 180
DB 121 DLFOATLDYGMYSREBELLRERKRIQTVGIASDYHOGSGTFLFOAGSGIYHVKDGPQG 180

DB 121 DLFOATLDYGMYSREBELLRERKRIQTVGIASDYHOGSGTFLFOAGSGIYHVKDGPQG 180

QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPMIAFISHSDIYISNIVTEEBRLTYVNEH 240

DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPMIAFISHSDIYISNIVTEEBRLTYVNEH 240

QY 241 ANMEDARSAGVATFVLQEEFDRYSYGMWCPKATTPSGGKILRLYEENDESEVEI 300

DB 241 ANMEDARSAGVATFVLQEEFDRYSYGMWCPKATTPSGGKILRLYEENDESEVEI 300

QY 301 TSPMLFTRRADSFYRPTGTANPKYTRMSGIMIDABGRIIDVIDKELIQPEEILFE 360

DB 301 TSPMLFTRRADSFYRPTGTANPKYTRMSGIMIDABGRIIDVIDKELIQPEEILFE 360

QY 361 YIARAGMTPEGKVAWSILLDRSOTRLOIVLISPELFIPEVDVNERORLISVPSYPL 420

DB 361 YIARAGMTPEGKVAWSILLDRSOTRLOIVLISPELFIPEVDVNERORLISVPSYPL 420

QY 421 IIVEETDIWINIHDFHVPQSHBEIEFIAPSECKTGFRHLKYTSILKESKYKSSG 480

DB 421 IIVEETDIWINIHDFHVPQSHBEIEFIAPSECKTGFRHLKYTSILKESKYKSSG 480

QY 481 GLPAPSPFKCPKKEINAITSGEWVLCRHGSIQVDEVRLVYEETGYDSPLEHLLYVS 540

DB 481 GLPAPSPFKCPKKEINAITSGEWVLCRHGSIQVDEVRLVYEETGYDSPLEHLLYVS 540

QY 541 YVNPGEVTRLDKGYSHSCCISQCHDFEISKYKQKPHCVSLKYLSSPEDDPTCKTEF 600

DB 541 YVNPGEVTRLDKGYSHSCCISQCHDFEISKYKQKPHCVSLKYLSSPEDDPTCKTEF 600

QY 601 WATILDSAGPLPDYTPPELIFSFESETGFTLYGMLYKPHDLOPKKYPVLFIYGGPVQL 660

DB 601 WATILDSAGPLPDYTPPELIFSFESETGFTLYGMLYKPHDLOPKKYPVLFIYGGPVQL 660

QY 661 VNNRFKGVKIFRLNTLASLGYVVVVIDNRGSCHRGLKPEGAFKYKMGQIEIDVVEGLQY 720

DB 661 VNNRFKGVKIFRLNTLASLGYVVVVIDNRGSCHRGLKPEGAFKYKMGQIEIDVVEGLQY 720

QY 721 LASRYDFIDDRVGIHGSYGYLSLMAIMQSRDIFRVAIAGAPVTLMIFYDTGYTERYM 780

DB 721 LASRYDFIDDRVGIHGSYGYLSLMAIMQSRDIFRVAIAGAPVTLMIFYDTGYTERYM 780

QY 781 GHPDNEQGYLGSVAMQAEKFPSEPNRLLLHGFLENVHFAHTSILSLFLVAKGKPYD 840

DB 781 GHPDNEQGYLGSVAMQAEKFPSEPNRLLLHGFLENVHFAHTSILSLFLVAKGKPYD 840

QY 841 LQIYPOERHSIRVPESGEHEHLHLHYLQENLGSRIALAKVI 882

DB 841 LQIYPOERHSIRVPESGEHEHLHLHYLQENLGSRIALAKVI 882

RESULT 8
ABB97362
ID ABB97362 standard; Protein; 724 AA.
XX
AC ABB97362;
XX
DT 27-JUN-2002 (first entry)
XX
DE Novel human protein SEQ ID NO: 630.
XX

KM Human; antihaemic; vulnerary; antiinflammatory; immunomodulator;
KM antifertility; cerebroprotective; cytosatic; rheumatic; gene therapy;
KM neuroprotective; antiparkinsonian; protein therapy; EST;
KM expressed sequence tag.

XX Homo sapiens.
XX
PN M0200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US26015.

```

XX 11-SEP-2000; 2000US-0659671.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Dimanac RT;
XX
XX WPI; 2002-292408/33.
DR
DR N-PSDB; AEN32548.
XX
PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX
XX Example 2; SEQ ID NO 630; 509pp; English.
PS
XX The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention.
XX
XX Sequence 724 AA;
SQ
Query Match 80.2%; Score 3771; DB 23; Length 724;
Best Local Similarity 82.1%; Pred. No. 0;
Matches 724; Conservative 0; Mismatches 0; Indels 158; Gaps 2;
QY 1 MAAAMETOLGVEIFETADCEENIESQDRPLPEFYERYWSQKLLADTRKHYGM 60
DB 1 MAAAMEEQLGVEIFETADCEENIESQDRPLPEFYERYWSQKLLADTRKHYGM 60
QY 61 AKAPHDPMFYKRNDDGPHSDRIYYLMSGENRENTLFYSEIPTKINRAAVLMSKPL 120
DB 61 AKAPHDPMFYKRNDDGPHSDRIYYLMSGENRENTLFYSEIPTKINRAAVLMSKPL 120
QY 121 DLPGATLDYGMVSEEEELRERKIGTVGIASVYHGSGGFLQAGSGIYHVDGPG 180
DB 121 DLPGATLDYGMVSEEEELRERKIGTVGIASVYHGSGGFLQAGSGIYHVDGPG 180
QY 121 DLPG----- 124
DB 121 DLPG----- 124
QY 181 FTQOPLRPNI VETS CPNIRMDPKLCPADPWIAFIHSNDIWSNI VTRERRLTYVANEL 240
DB 125 --QQPLRPNI VETS CPNIRMDPKLCPADPWIAFIHSNDIWSNI VTRERRLTYVANEL 182
QY 241 ANMEDARSAGVATFVYQEEFDYSGYWCPCAKETTPSGGKILILYEENDESEVEIIVH 300
DB 183 ANMEDARSAGVATFVYQEEFDYSGYWCPCAKETTPSGGKILILYEENDESEVEIIVH 242
QY 301 TSPMLERRRADSFPYPTKGTANPKVTFKMSIIMDAGRIIIVDKELIQPELLEFGVE 360
DB 243 TSPMLERRRADSFPYPTKGTANPKVTFKMSIIMDAGRIIIVDKELIQPELLEFGVE 302
QY 361 YIARAGTPEGKYAMSIILDRSQTRIQIVLISPFLIPVEDDVWERQRLIESVDSVTPL 420
DB 303 YIARAGTPEGKYAMSIILDRSQTRIQIVLISPFLIPVEDDVWERQRLIESVDSVTPL 362
QY 421 IIVEETDITNINHDIHFVFPQSHHEEIEFIASECTGGRHLYKTSIIKESYKXSSG 480
DB 363 IIVEETDITNINHDIHFVFPQSHHEEIEFIASECTGGRHLYKTSIIKESYKXSSG 422
QY 481 GLPAPSDFKPIKEEIAITSGEMVELGRHGSNIQVDEVRILVYEGTKDSPLEHHLVVS 540
DB 423 GLPAPSDFKPIKEEIAITSGEMVELGRHGSNIQVDEVRILVYEGTKDSPLEHHLVVS 482
QY 541 YVNPGEVTRLTDRGYSHSCISQHCDFPISKYSNOKNPHCVSLYKLSPPEDDPTCKTKEF 600
DB 483 YVNPGEVTRLTDRGYSHSCISQHCDFPISKYSNOKNPHCVSLYKLSPPEDDPTCKTKEF 542

```

```

QY 601 MATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIQPGKPYFTVLFIYGGPOVL 660
DB 543 MATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIQPGKPYFTVLFIYGGPOVL 599
QY 661 VNNRFKGVKFRNLNTASLGIVVVVINDRGSCHRGKFBGAFKXKQIEIDQVEGLQY 720
DB 600 ----- 599
QY 721 LASKYDIDLDVRGHHGWSYGVLSIMALMQRSDIFPVATAGAPVTLMIFPDGYTERYM 780
DB 600 -----VALAGAPVTLMIFPDGYTERYM 622
QY 781 GHPDQNEQGYVLSVAMQAEKFPSEPNRLILHGFIDENYFAHTSILSLFVRAKRPYD 840
DB 623 GHPDQNEQGYVLSVAMQAEKFPSEPNRLILHGFIDENYFAHTSILSLFVRAKRPYD 682
QY 841 LQIYPOERHSIRVSEGEHYELHLYLOENLSRIALKVI 882
DB 683 LQIYPOERHSIRVSEGEHYELHLYLOENLSRIALKVI 724

```

RESULT 9

ID ABG61600 standard; Protein; 658 AA.

AC ABG61600;

DT 12-AUG-2002 (first entry)

XX Human DPP-1 splice variant #7.

XX Human, serine protease; dipeptidyl peptidase IV-related protein; DPP; DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain; diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke; heart failure; hypertension; urinary retention; osteoporosis; cancer; ulcer; allergy; cancer; psychotic disorder; neurological disorder; dyskinnesia; reproductive disorder; inflammatory disorder; metabolic disorder.

OS Homo sapiens.

PN W0200231134-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.

PR 12-OCT-2000; 2000US-240117P.

XX (FERR) FERRING BV.

XX Q1 S, Akinsanya KO, Riviere PJ, Junien J;

DR WPI; 2002-444178/47.

DR N-PSDB; ABK83331.

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding the proteins, useful for treating e.g. fungal, bacterial, protozoan and viral infections, cancers, allergies, neurological disorders, or pain

XX Disclosure; Page 70-72; 113pp; English.

XX The present invention relates to the isolation of novel human serine proteases referred to as dipeptidyl peptidase IV (DPPIV)-related proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP) and nucleic acids encoding them are useful for treating infections such as fungal, bacterial, protozoan and viral infections, particularly infections caused by human immunodeficiency virus (HIV-1 or HIV-2), pain, diabetes, precocious puberty, infertility, obesity, anorexia, CC bulimia, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, angina pectoris, stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,

CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

XX Sequence 658 AA;

Query Match 74.6%; Score 3504; DB 23; Length 658;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAAAMETQQLGVEIFETADCEENTESQDRPKLEPFYVERYSQKLLADTRKHGYM 60
DB 1 MAAAMETQQLGVEIFETADCEENTESQDRPKLEPFYVERYSQKLLADTRKHGYM 60
QY 61 AKAPHEMFVGRNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTKINRAAVLMSKPELL 120
DB 61 AKAPHEMFVGRNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTKINRAAVLMSKPELL 120
QY 121 DLFOATLDYGMYSREBEILLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXDGPGQG 180
DB 121 DLFOATLDYGMYSREBEILLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXDGPGQG 180
QY 121 DLFOATLDYGMYSREBEILLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXDGPGQG 180
DB 121 DLFOATLDYGMYSREBEILLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXDGPGQG 180
QY 181 FTQOPLRPNLVETSGPNIRMDPKLCPADPDWIAFIHSDNIMSNIVTEEBERLLTYVHNEL 240
DB 181 FTQOPLRPNLVETSGPNIRMDPKLCPADPDWIAFIHSDNIMSNIVTEEBERLLTYVHNEL 240
QY 241 ANMEDARSAGVATFVLOEEDFRYSGYWMCCKAETTPSGKILRLIYEENDESEVEIITHV 300
DB 241 ANMEDARSAGVATFVLOEEDFRYSGYWMCCKAETTPSGKILRLIYEENDESEVEIITHV 300
QY 301 TSPMLERTRADSFRYPKGTANPKVTFKMSIEMIDAGR11DIVDKELIQPEILFEGVE 360
DB 301 TSPMLERTRADSFRYPKGTANPKVTFKMSIEMIDAGR11DIVDKELIQPEILFEGVE 360
QY 361 YIARAGWTPBEKAYMSILLDRSQRLQIVLISPELFIPEDDVWERQRLISVDSVTPPL 420
DB 361 YIARAGWTPBEKAYMSILLDRSQRLQIVLISPELFIPEDDVWERQRLISVDSVTPPL 420
QY 421 IIVETTDIMINIHDIHFVPOSHSEIEFTFASECKTFPHLYKITSILKESKYSRSG 480
DB 421 IIVETTDIMINIHDIHFVPOSHSEIEFTFASECKTFPHLYKITSILKESKYSRSG 480
QY 481 GLPAPSDFKPIKEIAITTSGBEVLGHSNIOYDEVRLVYFEGTKDSPLEHNLVYVS 540
DB 481 GLPAPSDFKPIKEIAITTSGBEVLGHSNIOYDEVRLVYFEGTKDSPLEHNLVYVS 540
QY 541 YVNGEVTURLTDRGSHSCISOHCDPFISKYSNOKNHCYSIKLSSPEDDPCTCKTEF 600
DB 541 YVNGEVTURLTDRGSHSCISOHCDPFISKYSNOKNHCYSIKLSSPEDDPCTCKTEF 600
QY 601 WATILDSAGPLPDYTPPEIFSESTTGTLYGMLYKPHDLQPGKKYPTVLEFYGG 655
DB 601 WATILDSAGPLPDYTPPEIFSESTTGTLYGMLYKPHDLQPGKKYPTVLEFYGG 655

```

RESULT 10
 ABG61596
 ID ABG61596 standard; Protein; 661 AA.
 XX
 AC ABG61596;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPRP-1 splice variant #3.
 XX

KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinesia; reproductive disorder; inflammatory disorder;
 KM metabolic disorder.

XX Homo sapiens.
 OS
 XX
 PN W0200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 XX
 DR WPI; 2002-444178/47.
 XX
 DR N-P8DB; ARK83327.
 XX

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -

PS Disclosure; Page 63-65; 113pp; English.

CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

SQ Sequence 661 AA;

Query Match 74.6%; Score 3504; DB 23; Length 661;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAAAMETQQLGVEIFETADCEENTESQDRPKLEPFYVERYSQKLLADTRKHGYM 60
DB 1 MAAAMETQQLGVEIFETADCEENTESQDRPKLEPFYVERYSQKLLADTRKHGYM 60
QY 61 AKAPHEMFVGRNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTKINRAAVLMSKPELL 120
DB 61 AKAPHEMFVGRNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTKINRAAVLMSKPELL 120
QY 121 DLFOATLDYGMYSREBEILLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXDGPGQG 180
DB 121 DLFOATLDYGMYSREBEILLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXDGPGQG 180
QY 121 DLFOATLDYGMYSREBEILLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXDGPGQG 180
DB 121 DLFOATLDYGMYSREBEILLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXDGPGQG 180
QY 181 FTQOPLRPNLVETSGPNIRMDPKLCPADPDWIAFIHSDNIMSNIVTEEBERLLTYVHNEL 240
DB 181 FTQOPLRPNLVETSGPNIRMDPKLCPADPDWIAFIHSDNIMSNIVTEEBERLLTYVHNEL 240
QY 241 ANMEDARSAGVATFVLOEEDFRYSGYWMCCKAETTPSGKILRLIYEENDESEVEIITHV 300
DB 241 ANMEDARSAGVATFVLOEEDFRYSGYWMCCKAETTPSGKILRLIYEENDESEVEIITHV 300
QY 301 TSPMLERTRADSFRYPKGTANPKVTFKMSIEMIDAGR11DIVDKELIQPEILFEGVE 360
DB 301 TSPMLERTRADSFRYPKGTANPKVTFKMSIEMIDAGR11DIVDKELIQPEILFEGVE 360
QY 361 YIARAGWTPBEKAYMSILLDRSQRLQIVLISPELFIPEDDVWERQRLISVDSVTPPL 420
DB 361 YIARAGWTPBEKAYMSILLDRSQRLQIVLISPELFIPEDDVWERQRLISVDSVTPPL 420

```


QY 421 IYEEETDIIWIMINIHDFHVPFQSHHEEIEFIFASECKTGFRHLKYKITSILKSKYKRSSG 480
 DB 421 IYEEETDIIWIMINIHDFHVPFQSHHEEIEFIFASECKTGFRHLKYKITSILKSKYKRSSG 480
 QY 481 GLPAPSDFKCP1KEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVS 540
 DB 481 GLPAPSDFKCP1KEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVS 540
 QY 541 YVNPGEVTRLTDGSGSHSCCISOHCDPFISKYSNOKNPHCVSLVYLSSPEDDPTCKTERF 600
 DB 541 YVNPGEVTRLTDGSGSHSCCISOHCDPFISKYSNOKNPHCVSLVYLSSPEDDPTCKTERF 600
 QY 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGG 655
 DB 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGG 655

RESULT 11

ID ABG61594 standard; Protein; 690 AA.

AC ABG61594;

DT 12-AUG-2002 (first entry)

DE Human DPRP-1 splice variant #1.

KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KM DPPIV; infection; human immunodeficiency virus; HIV-1, HIV-2, pain;
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinetia; reproductive disorder; inflammatory disorder;
 KM metabolic disorder.

OS Homo sapiens.

EN WO200231134-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.

PR 12-OCT-2000; 2000US-240117P.

PA (FERR) FERRING BV.

PI Qi S, Akinsanya KO, Riviere PJ, Junien J;

DR WPI, 2002-444178/47.

DR N-PEDB; ABR83325.

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain

PS Disclosure; Page 59-61; 113pp; English.

CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and

CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

XX Sequence 690 AA:

Query Match 74.6%; Score 3504; DB 23; Length 690;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAAMETQOLGVEIFETADCEENIESODRPKLEPPEYVERYSQKLLADTRKXGVM 60

DB 1 MAAAMETQOLGVEIFETADCEENIESODRPKLEPPEYVERYSQKLLADTRKXGVM 60

QY 61 AKAPHDFFVVRKNDPDGHSRDIYVLAAMSGENRENTLFYSELPKTIINRAVAMLWKPL 120

DB 61 AKAPHDFFVVRKNDPDGHSRDIYVLAAMSGENRENTLFYSELPKTIINRAVAMLWKPL 120

QY 121 DLFQATLDYGMYSREBELRERKIGTVGIASVDYHQSGLFLPQAGSIVYVKQGPQG 180

DB 121 DLFQATLDYGMYSREBELRERKIGTVGIASVDYHQSGLFLPQAGSIVYVKQGPQG 180

QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTEERRLTVVHNL 240

DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTEERRLTVVHNL 240

QY 241 ANMEEDARSAGVATFVLQEEFDRYSGYWMCPEAETTPESGKILRLYEENDESEVEI 300

DB 241 ANMEEDARSAGVATFVLQEEFDRYSGYWMCPEAETTPESGKILRLYEENDESEVEI 300

QY 301 TSPMLFTRRAASFRRPKGTANPKTFKSGEIMDABRIIDVIDKELOPELIFEGVE 360

DB 301 TSPMLFTRRAASFRRPKGTANPKTFKSGEIMDABRIIDVIDKELOPELIFEGVE 360

QY 361 YIARAGWPEKGYAMSIILDRSQTRLQVLSPELFIVEDVWERQRLIEVPSVPL 420

DB 361 YIARAGWPEKGYAMSIILDRSQTRLQVLSPELFIVEDVWERQRLIEVPSVPL 420

QY 421 IYEEETDIIWIMINIHDFHVPFQSHHEEIEFIFASECKTGFRHLKYKITSILKSKYKRSSG 480

DB 421 IYEEETDIIWIMINIHDFHVPFQSHHEEIEFIFASECKTGFRHLKYKITSILKSKYKRSSG 480

QY 481 GLPAPSDFKCP1KEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVS 540

DB 481 GLPAPSDFKCP1KEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVS 540

QY 541 YVNPGEVTRLTDGSGSHSCCISOHCDPFISKYSNOKNPHCVSLVYLSSPEDDPTCKTERF 600

DB 541 YVNPGEVTRLTDGSGSHSCCISOHCDPFISKYSNOKNPHCVSLVYLSSPEDDPTCKTERF 600

QY 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGG 655

DB 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGG 655

RESULT 12

ID AAB93565 standard; Protein; 632 AA.

AC AAB93565;

DT 26-JUN-2001 (#first entry)

DE Human protein sequence SEQ ID NO:12964.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ Sequence 613 AA;

Query Match 68.9%; Score 3236; DB 23; Length 613;
Best Local Similarity 100.0%; Pred. No. 4, 2e-303;
Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSQKLKLLADTRKHYGM 60
DB 1 MAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSQKLKLLADTRKHYGM 60
QY 61 AQAPEHFMFKKNDPDGPHSDRIYIAMSGENRENTLFYSEIPKTNRAAVLMLSWKPLL 120
DB 61 AQAPEHFMFKKNDPDGPHSDRIYIAMSGENRENTLFYSEIPKTNRAAVLMLSWKPLL 120
QY 121 DLFQATLDYGMYSREBELLRERKRTGTGAGSYDHQSGTFLFQSGGIYHVKOGPG 180
DB 121 DLFQATLDYGMYSREBELLRERKRTGTGAGSYDHQSGTFLFQSGGIYHVKOGPG 180
QY 181 FTQOPLRPNLVETSCENIMDPKLCPADPDWIAFIHNDIWTISNTYREERLTYVHNEL 240
DB 181 FTQOPLRPNLVETSCENIMDPKLCPADPDWIAFIHNDIWTISNTYREERLTYVHNEL 240
QY 241 AMMEEDARAGVATFLQSEFDRISYMWCPKAEITPSSGKILRIIYENDSEVEIHHV 300
DB 241 AMMEEDARAGVATFLQSEFDRISYMWCPKAEITPSSGKILRIIYENDSEVEIHHV 300
QY 301 TSPMLETRADSEFRYPKGTANPKYFKMSEIMIAEGSIIDIVIKELQPEILLFEGVE 360
DB 301 TSPMLETRADSEFRYPKGTANPKYFKMSEIMIAEGSIIDIVIKELQPEILLFEGVE 360
QY 361 YIARAGWPEEGKXAMSIILDRSQTRLQIVLISPELFIPEYDDVMERQRLIEVPSVPL 420
DB 361 YIARAGWPEEGKXAMSIILDRSQTRLQIVLISPELFIPEYDDVMERQRLIEVPSVPL 420
QY 421 ILYETTDIWINIHDIHFVFPQSHREIRFIASCKYGRHLYKITSLKSKYKRSRG 480
DB 421 ILYETTDIWINIHDIHFVFPQSHREIRFIASCKYGRHLYKITSLKSKYKRSRG 480
QY 481 GLPAPSDFCPIKEELIATISGEMVLAGHGSNIQYDEVRLVYFPGTQSPLEHLLYVVS 540
DB 481 GLPAPSDFCPIKEELIATISGEMVLAGHGSNIQYDEVRLVYFPGTQSPLEHLLYVVS 540
QY 541 YVNPBEVTRLTDKRGYSHSCCISQHCDFISKYSNQKNPHCVSLYKLSBEDDPTCKTEF 600
DB 541 YVNPBEVTRLTDKRGYSHSCCISQHCDFISKYSNQKNPHCVSLYKLSBEDDPTCKTEF 600
QY 601 WATIIIDS 607
DB 601 WATIIIDS 607

RESULT 14
ID ABG61592 standard; Protein; 863 AA.
XX ABG61592;
XX 12-AUG-2002 (first entry)
XX Human DPRP related serine protease DPRP-2.
XX Human, serine protease, dipeptidyl peptidase IV-related protein, DPRP,
XX DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
XX diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;

KM heart failure, hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM dyskinesia; reproductive disorder; inflammatory disorder;
KM metabolic disorder.

XX Homo sapiens.

PN MO200231134-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.

PR 12-OCT-2000; 2000US-240117P.

PA (FERR) FERRING BV.

PI Qi S, Akinsanya KO, Riviere PJ, Junien J;

DR WPI; 2002-444178/47.

DR N-PSDB; ABK63323.

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain

PS Claim 17; Fig 1; 113pp; English.

CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

SQ Sequence 863 AA;

Query Match 61.1%; Score 2870; DB 23; Length 863;
Best Local Similarity 61.5%; Pred. No. 1.8e-267;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;

QY 35 FVYERYSQKLKLLADTRKHYGMMAKAPDFMFVKKNDPDGPHSDRIYIAMSGENE 94
DB 35 FVYERYSQKLKLLADTRKHYGMMAKAPDFMFVKKNDPDGPHSDRIYIAMSGENE 94
QY 24 FVQVGHSDGLRSITIHGSRKISGLIVKAPDFMFVQKTDSEGPSHRLVYGMFYGSSE 83
DB 24 FVQVGHSDGLRSITIHGSRKISGLIVKAPDFMFVQKTDSEGPSHRLVYGMFYGSSE 83
QY 95 NTLFYSEIPKTNRAAVLMLSWKPLDLFOATLDYGMYSREBELLRERKRTGTGAGSYD 154
DB 95 NTLFYSEIPKTNRAAVLMLSWKPLDLFOATLDYGMYSREBELLRERKRTGTGAGSYD 154
QY 84 NSLIYSEIPKTVKREKALLLSWKQMLDHFQATLPHNGVYSREBELLRERKRLGVFETIYD 143
DB 84 NSLIYSEIPKTVKREKALLLSWKQMLDHFQATLPHNGVYSREBELLRERKRLGVFETIYD 143
QY 155 YHOGSGTFLFOAGSGCIYHVKGGFQGFQOPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214
DB 155 YHOGSGTFLFOAGSGCIYHVKGGFQGFQOPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214
QY 144 FHSBSGLFLFOASNSLFCRCDGNGGFVVSFMKYLEIKTQCSGPRMDKICPADPAFPSF 203
DB 144 FHSBSGLFLFOASNSLFCRCDGNGGFVVSFMKYLEIKTQCSGPRMDKICPADPAFPSF 203
QY 215 IHSNDIWTISNTYREERLTYVHNELAMMEEDARAGVATFLQSEFDRISYMWCPKAE 274
DB 215 IHSNDIWTISNTYREERLTYVHNELAMMEEDARAGVATFLQSEFDRISYMWCPKAE 274
QY 204 INNSDIWAMANTETBERRLTCHQGLSNVLDPPKSAGVATVIOBEFDRFTGYMWCPYAS 263
DB 204 INNSDIWAMANTETBERRLTCHQGLSNVLDPPKSAGVATVIOBEFDRFTGYMWCPYAS 263
QY 275 TTPSGG-KILRIIYENDSEVEIHHVTSFPMLETRRADSEFRYPKGTANPKYFKMSEIM 333
DB 275 TTPSGG-KILRIIYENDSEVEIHHVTSFPMLETRRADSEFRYPKGTANPKYFKMSEIM 333
QY 264 WEGSEGKTLRIILYEEVDESEVEIHHVSPALERKTDYVYPTGSGNPKIALKLAFFQ 323
DB 264 WEGSEGKTLRIILYEEVDESEVEIHHVSPALERKTDYVYPTGSGNPKIALKLAFFQ 323
QY 336 IDARGRIIDVIDEKLQPELLEFGVYIARAGWPEEGKXAMSIILDRSQTRLQIVLISP 393
DB 336 IDARGRIIDVIDEKLQPELLEFGVYIARAGWPEEGKXAMSIILDRSQTRLQIVLISP 393
QY 324 TDSQGIKIVSQEKLVPFSSLFEPKVEYIARAGWTRDGKXAMAMFLDRPQOOLQIVLLP 383
DB 324 TDSQGIKIVSQEKLVPFSSLFEPKVEYIARAGWTRDGKXAMAMFLDRPQOOLQIVLLP 383

QY	39	ELIPIPEDVMEQRIIEBPSVMPPLIIESTTDIWININDIFHVPQSH-REEIEFIF	452
Db	384	ALFIIPSTENEOQLASARAVPRNOPYVVEEVTWIMVNDIIRYPPQSGEDELFLR	443
QY	453	ASECKTGFRHLTKITSILKEKYYKSSGGI.PAPSDFKCPRIKEEIALTNGEMEVLGRGNS	512
Db	444	ANBECKTGFCMLYKVTAVLKSQGYDMSEPSPEDEBFCPIKEEIALTNGEMEVLARGSK	503
QY	513	IQDVEVRLIYFEFGTDSPLREHLLIYVSVVNGEYTRLTDBGYSHSCISIQHCFPIISKY	572
Db	504	IWNEELTKLIVFGQIDTPLREHLLIYVSVYEAAGELVRLTTPGFSHSCSMQSNFMFMFVSHY	563
QY	573	SNOKNPHCVSLIYLSPPDDPTCKTKEFATLIDAGELPPTYPELFSFESTGFTLYG	632
Db	564	SSVSTPCCVHYLTSGPDDDLHKQRFASMMWEASAECRPDYVPEIFHFTTRSDVRLYG	623
QY	633	MLYKPHDLQPKKYPVLEFIYGGPOVLVNNRFKGVKFRILNTLASLGIVVVVITDNGSC	652
Db	624	MIYKPHALQGGKHPLVLEFVGGPOVLVNNRSFKIKYLRILNTLASLGIVVVITDNGRSC	663
QY	693	HRGLKFECAKRYKMGQIEIIDQVEGLQYULASRYDPLDRGIGHMSGGVLSMALMOR	752
Db	684	QRGIRFEFGALKMOMGOVEIEDQVEGLQFAEAKTGTIDLSRAIHMSTYGGSLMGLIHK	743
QY	753	SDIFRVAIAGAPVTLMIFYDTGYTERMYGHPDQNEQGYLGSVAMQAEKPSPEENRLLIL	812
Db	744	POVFKAVALIAGAPVTVMMAVDGTGYTERYMDVPENNQHGAEASVALMHWKLPENENRLLIL	803
QY	813	HGFLDENVVAHNSILSLFLVAKGRYDQIYPOERHSIRVPESGEHEMLLHYLOENL	872
Db	804	HGFLDENVVEFHNNELVSQLIRAGKGYQOQIYPNRSHSIRCEGSEHEVTLHFLFOEYL	863

RESULT 15	ABR98134	standard; Protein; 892 AA.
XX	ABR98134	
AC	ABR98134;	
XX		
DT	17-OCT-2002	(first entry)
XX		
DE	Human PMW Incyte ID 7972712CD1.	
XX		
KM	Human; PMW; protein modification and maintenance molecule;	
KM	anticonvulsant; neuroprotective; nootropic; cytosolic; antipsoriatic;	
KM	antianesthetic; dermatological; antidiabetic; antiparkinsonian;	
KM	antianemic; antiinflammatory; antitumor; antitubercular; cardiant;	
KM	hepatocytic; osteopathic; antileptic; antipyretic; virucide;	
KM	antibacterial; fungicidal; gastroenteric; antidiarrhetic; laxative;	
KM	haemostatic; anti-HIV; thrombolytic; anticoagulant; gynaecological;	
KM	cardiovascular; antihypertensive; hypotensive; vasotropic;	
KM	antitumor; antirheumatic; immunosuppressive; antiallergic; antithyroid;	
KM	nephrotoxic; antipain; thyromimetic; antirheumatic; uteropic;	
KM	ophthalmological; antiparasitic; tranquilizer; vulnerary; keratolytic;	
KM	auditory; antiseborrheic; antidepressant; neuroleptic; antinfertility;	
KM	antileptic; procoagulant; Crohn's disease; hypertension;	
KM	autoimmune; inflammatory; anaemia; cell proliferative; developmental;	
KM	epithelial; scabies; neurological; Alzheimer's disease; reproductive;	
KM	ectopic pregnancy; gene therapy; vaccine; disorder;	
KM	dipeptidyl peptidase 8.	
XX		
XX	Homo sapiens.	
XX		
XX	MO200246383-A2.	
PN		
PD	13-JUN-2002.	
XX		
PF	05-DEC-2001; 2001MO-US46964.	
XX		
XX	08-DEC-2000; 2000US-254399P.	
PR	21-DEC-2000; 2000US-257803P.	
RR	05-JAN-2001; 2001US-260110P.	

PR 19-JAN-2001; 2001US-262851P.
PR 25-JAN-2001; 2001US-264623P.

(INCY-) INCYTE GENOMICS INC.

P1 Yue H., Azimzai Y., Kallick DA., Baughn MR., Griffin JA., Swarnakar A;
P1 Lai PQ., Walia NK., Hafalia ADA., Gandhi AR., Au-Young J., Elliott VS;
P1 Ramkumar J., Thangavelu K., Lu Y., Warren BA., Lu DMW., Lee EA,
P1 Tribouley CM., Arvizu C., Delegeane AM., Yao MG., Khan FA;
P1 Sanjanwala M;

XX WPI; 2002-519664/55.
DR N-PDSB; ABQ75955.

XX
XX
XX New isolated Protein Modification and Maintenance polypeptides; useful
PT for diagnosis, and treatment of e.g. gastrointestinal disorders -
PS Claim 1 (a); Page 172-174; 200DP; English.

The invention relates to an isolated protein modification and maintenance
CC (PMW) polypeptide. Polypeptides of the invention may be used in the
CC diagnosis, treatment and prevention of disorders associated with
CC decreased expression or activity of PMW. These include gastrointestinal
disorders (e.g. Crohn's disease), cardiovascular disorders (e.g.
hyperextension), autoimmune/inflammatory disorders (e.g. anaemia), cell
proliferative disorders, developmental disorders, epithelial disorders
CC (e.g. psoriasis), neurological disorders (e.g. Alzheimer's disease)
reproductive disorders (e.g. ectopic pregnancy), and in gene therapy or a
vaccine for such diseases. They may also be used in the assessment of the
effects of exogenous compound on the expression of nucleic acid and amino
acid sequences of protein modification and maintenance molecules. The
current sequence represents a human PMW of the invention, which has been
found to have homology with human dipeptidyl peptidase 8.

Sequence 892 AA:

Query Match# 61.1%; Score 2870; DB 23; Length 892;
Best Local Similarity 61.5%; Pred. No. 1.9e-267;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2

D# 35 FLYERSNWSQLKLADLRKHGNYMAAPRDMFVKRNBDPDGHSRTIYUAMSGENE 94
|||:::||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
D# 53 FOVQRKSWGLSLIIHSGRKYSGLIVNAAPPDFOVKTBSBGSHRLYLLOMPYGSR 112
NTLFYEIRPKTINRAAVLMLSMKPLDLDFOATLDYGMVSEBELRRKKIGTVGIASYD 154
113 NSLLVEIRIKKYKRKAALLLSMKMDLBFQATRHNGVSYREBLRRKKLGFGITSD 172

Q# 155 YHGSGSTFLFOAGSGIYNVGOGROGTQORPLRNLVETSCPNIRNDPKICPADPWIAF 214
FHSESGLPFQASNSLFFCRDGKGKGFVSFMKLELTKTCOSGRMDPKICPADPAFFSF 232
173

D# 215 ITHSNDWININTREERELTYVNEHLAAMEDASAAVAFTVOEEDRSYGVMCPKE 274
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
D# 233 INMSDLMTVANITGERKRLTRCHQSLNVADDPPSAAGALFYIOEEDRDRTGWMCPTLS 292

Q# 275 TTPSGG-KILRIIYENDSEVEIITHVTSPMLETRRADSRYPKGTANKPTFKXSELIM 333

D# 293 WEGBESGLKTRLITYEVDESEVENIHVPSPALERKTDTSRYRPRTGSKNPKIALKLAERO 352

Q# 334 IDAGEGIIDVIDKEILOPEELLPEGUYERIAAGCTPGSKAKMGLILDROSTRLOIYLISP 393
||:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
D# 353 TSSQGKIYSTOKELIVOPSSLFPVEVERIAAGCTRGGKAYMMFDRPOOWLOVLDDP 412

Q# 394 ELTIPEEDDVOMEKRLIESVDSTPLITYETTIDIWINIHDIFEHVPOSH-EETIEFTIP 452
||'':||:'':||:'':||:'':||:'':||:'':||:'':||:'':||:'':||:'':||:
D# 413 ALFIPTENEBEORLASARAARVRNNQPVYYVEEVTNWNINHDIIFYRPOSGBEGDELCLFR 472

Q# 453 ASECKTGPHLIYTITSILKSKRYRSSGGLPAPDEFKCPIKEBIATITSGEWLVLAGHGN 512
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
D# 473 ANECKGFCHELTXYKNAVIVKSQGYDWMBSEFSGDEGEFCPIKEBIALTITSGEMWTARHGSK 532

Q# 513 IQVDEVBRILVYEEGKDPLEBNHYVYSYVNPAGVRLITDRGVSSHSCCISOHQDDFITSX 572

```

Db      533  I W N E E T K L V Y F O G T K O T P L E H H L Y V V S Y E A G E I V R L T P G F S H S C S M S Q N F D M F V S H Y 592
Qy      573  S N O K N P C V S L Y K L S S P E D D P T C K T K E F W A T I I D S A G E L P Y T P P E I F S S E S T G F T L Y G 632
Db      593  S S V S T P C V A H Y K L S G P D D D P L H K O P R F W A S M E A S C P D Y V P P E I F H F H T R S D V R L Y G 652
Qy      633  M L Y K P H D L O P G K K Y P T V L F I Y G P O V O L V N N R F K G V K Y F R L N T L A S L G Y V V V I D N R G S C 692
Db      653  M I Y K P H A L O P G K K H P T V L F V Y G P O V O L V N N S P F K I K Y L R I N T L A S L G Y A V V I D G R G S C 712
Qy      693  H R G L K E R G A F K Y K M G O I E I D D O V E G L O Y L A S R Y D F I D L D R V G I H G M S Y G Y L S I M A L M O R 752
Db      713  Q R G L R F E G A L K N O M G O V E I E D O V E G L O F V A E K Y F I D L S R V A I H G M S Y G F L S I M G L I H K 772
Qy      753  S D I F R V A I A G A P V T L M I F Y D T G Y T E R Y W G H P D N O G Y L G S V A M O A E K P S E P N R L L L 812
Db      773  P O V F K V A I A G A P V T V M A Y D I G Y T E R M D V P E N N O H G Y E A G S V A L H V E K L P N E P N R L L I L 832
Qy      813  H G F L D E N V H F A H T S I L S F L V R A G K P Y D L Q I Y P O E R H S I R V P E S G E H Y E L H L L A Y L O E N L 872
Db      833  H G F L D E N V H F H T N F L V S O L I R A G K P Y O L Q I Y P N E R H S I R C P E S G E H Y E V T L I H F L Q E Y L 892

```

Search completed: October 15, 2003, 17:10:43
 Job time : 58.7856 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:10:56 ; Search time 19.6777 Seconds
(without alignments)
1896.467 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700
Sequence: 1 MAAMETEQGVGEIFETAD.....HLHYLOENLSRIALKVI 882

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/1aa/5A COMB.pdp.*
2: /cgn2_6/ptodata/1/1aa/5B COMB.pdp.*
3: /cgn2_6/ptodata/1/1aa/6A COMB.pdp.*
4: /cgn2_6/ptodata/1/1aa/6B COMB.pdp.*
5: /cgn2_6/ptodata/1/1aa/PTCUS COMB.pdp.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1645.5	35.0	310	4	US-09-794-236-4
2	582.5	12.4	771	4	US-09-462-284-2
3	529	11.3	766	5	US-10-002-593-6
4	525	11.2	755	5	PCT-US93-07923-3
5	525	11.2	759	5	PCT-US93-07923-2
6	525	11.2	766	1	US-08-230-491A-3
7	525	11.2	766	1	US-08-619-280A-3
8	525	11.2	766	2	US-08-940-391-3
9	525	11.2	766	4	US-09-794-236-1
10	454	9.7	760	1	US-08-230-491A-2
11	454	9.7	760	1	US-08-619-280A-2
12	454	9.7	760	2	US-08-940-391-2
13	282	6.0	632	3	US-09-016-080-1
14	249.5	5.3	657	4	US-09-355-166-1
15	225.5	4.8	593	5	PCT-US93-07923-11
16	192	4.1	622	2	US-08-664-646A-2
17	192	4.1	622	2	US-09-066-285-2
18	192	4.1	622	3	US-09-261-006-2
19	182	4.1	622	3	US-08-951-088-2
20	182	4.1	622	3	US-09-566-2
21	192	4.1	622	4	US-09-609-570-2
22	192	4.1	622	4	US-09-427-372-2
23	192	4.1	622	4	US-09-693-554-2
24	180.5	3.8	622	4	US-09-390-234-20
25	175.5	3.7	614	4	US-09-252-991A-20060
26	133	2.8	344	4	US-09-724-623-79
27	132.5	2.8	344	4	US-09-198-452A-181

28	123	2.6	1022	1	US-08-271-364A-8	Sequence 8, Appl
29	123	2.6	1022	2	US-08-222-715B-27	Sequence 27, Appl
30	120.5	2.6	686	3	US-09-368-169-8	Sequence 8, Appl
31	116	2.5	561	3	US-09-233-989-5	Sequence 5, Appl
32	113.5	2.4	1209	4	US-09-749-588-4	Sequence 4, Appl
33	112.5	2.4	529	4	US-09-252-991A-24711	Sequence 24711, A
34	112.5	2.4	1285	1	US-07-582-945-2	Sequence 2, Appl
35	112.5	2.4	1285	2	US-08-453-141-2	Sequence 2, Appl
36	112.5	2.4	1285	2	US-08-293-314-2	Sequence 2, Appl
37	110.5	2.4	836	4	US-09-491-356C-21	Sequence 21, Appl
38	110	2.3	1296	1	US-08-480-604A-28	Sequence 28, Appl
39	110	2.3	1296	2	US-08-405-496A-28	Sequence 28, Appl
40	110	2.3	1296	3	US-08-915-136-28	Sequence 28, Appl
41	109	2.3	570	3	US-09-068-960-13	Sequence 13, Appl
42	109	2.3	808	4	US-09-198-452A-327	Sequence 327, App
43	108.5	2.3	782	4	US-09-585-858-29	Sequence 29, Appl
44	106	2.3	256	4	US-09-355-166-2	Sequence 2, Appl
45	106	2.3	1362	4	US-09-134-001C-4955	Sequence 4955, Ap

ALIGNMENTS

RESULT 1
US-09-794-236-4
Sequence 4, Application US/09794236
Patent No. 6337069
GENERAL INFORMATION:
APPLICANT: Grouzmann, Eric
APPLICANT: Lacroix, Jean-Silvain
APPLICANT: Monod, Michel
TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
FILE REFERENCE: 81985/276823
CURRENT APPLICATION NUMBER: US/09/794,236
CURRENT FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 310
TYPE: PRT
ORGANISM: Homo sapiens
US-09-794-236-4

Query Match 35.0%; Score 1645.5; DB 4; Length 310;
Best Local Similarity 86.4%; Pred. No. 5.2e-159;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

QY	524	FEETKDSPLHHLVYVSYNPGVYTRITDRGYSHSCISQHCDFEISKYSNOKNPHCVSL	583
DB	1	FEETKDSPLHHLVYVSYNPGVYTRITDRGYSHSCISQHCDFEISKYSNOKNPHCVSL	60
QY	584	YKLSPPDDPTCKTKERFWATILDSAGPLDPTTRPEIFSEFSTTGTLYGMLYKPHDLOPG	643
DB	61	YKLSPPDDPTCKTKERFWATILDSAGPLDPTTRPEIFSEFSTTGTLYGMLYKPHDLOPG	120
QY	644	KKYPVLYFYGGPOVOLLVNNRFKGVKXFRNLATLSLGYVVVVDNRSGHGLKFECAF	703
DB	121	KKYPVLYFYGGPOVOLLVNNRFKGVKXFRNLATLSLGYVVVVDNRSGHGLKFECAF	134
QY	704	YKRGQLEIDQVGVGLGYLARSYDFIDLDRVINGWSYGYLSLMAIMORSIDIFRVALAGA	763
DB	135	YKRGQLEIDQVGVGLGYLARSYDFIDLDRVINGWSYGYLSLMAIMORSIDIFRVALAGA	191
QY	764	PVTLMTFYDTGYTERYMGHPDQVGGYVLSVAMQEKPFSENRLLHLHGFDENVHFA	823
DB	192	PVTLMTFYDTGYTERYMGHPDQVGGYVLSVAMQEKPFSENRLLHLHGFDENVHFA	251
QY	824	HTSILLSFLVRAGKPYDLOIYPOERHSIRVPSGEHYELHLHYLOENLSRIALKVI	882
DB	252	HTSILLSFLVRAGKPYDLOIYPOERHSIRVPSGEHYELHLHYLOENLSRIALKVI	310

RESULT 2

US-09-462-284-2
 ; Sequence 2, Application US/09462284
 ; Patent No. 6309868
 ; GENERAL INFORMATION:
 ; APPLICANT: Nestec S.A.
 ; APPLICANT: Monod, Michel
 ; APPLICANT: Donnas, Agnes
 ; APPLICANT: Affolter, Michael
 ; APPLICANT: Van Den Broek, Peter
 ; TITLE OF INVENTION: CLONING OF THE
 ; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
 ; FILE REFERENCE: 8265-298
 ; CURRENT APPLICATION NUMBER: US/09/462,284
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 771
 ; TYPE: PRT
 ; ORGANISM: Fungus
 ; US-09-462-284-2

Query Match 12.4%; Score 582.5; DB 4; Length 771;
 Best Local Similarity 28.4%; Pred. No. 7,6e-50;
 Matches 198; Conservative 103; Mismatches 262; Indels 135; Gaps 30;
 QY 203 KLCRADP-----W-----IAFHSND--IWISNIYREERLLTYVHNELANMEE 245
 DB 136 KSVLMPQEGSDIOYAQMSPVGNITIAFVRENDLYVWDGTVVR-----ITD 181
 QY 246 DA---RSAGVATFVLQEEF--DRYGYWMCRAETTPSGKILRIIYEENDESEVEIITHV 300
 DB 182 DGGDFMHGVDPMIYEEIILDRYA-LMFSPDE-----YIAYISFNETGPTVTV 231
 QY 301 TSPMLERRADSF-----RYPKGTANPKVTFKXSEIMDAEGRIIDVIDIKELIOPFEI 354
 DB 232 QYVMDNOEIAPAYMELKIRYPKVSQTPVTLS-----LNIATKEVQAID 280
 QY 355 LFEVEYIARAGMPEEGVAMSIILDRSQTLQIVLISPELPIYVEDDVMERORLISVP 414
 DB 281 AFESTDLIT-----GEVAM--LTDHTT-----VAAKAFNRVOD--QOKVAVD7AS 323
 QY 415 DSVPLIYEETDIWI-NIHDIHFVFPQSHHEIEFIASECKTGPHLYKISILKES 473
 DB 324 NKAT-VISDRGDTGMDNLMSKYGIPKSDKDAYIIDISGMAHLYLP----- 376
 QY 474 KYRSGGGLPAPSDFKCPKEEIAITSGEMEVILGRHSNIOQDEVRLVYFEGTKDSPLE 533
 DB 377 ---VSGGEPIP-----LTKGDMEVT---SILSDQERQLVYVYLSQHHSTE 416
 QY 534 HHLVYVYVNPGEVYTRITDRGYS--HSCCISQHCDFISKYSNOKNHCVSILKSPED 591
 DB 417 RHLXSVY-STFAVTPLVYDITVAAYMSASFSANGYVILTYGGPDVY-DELVTNS--- 471
 QY 592 DPTCKTEFWATILDSAGPL---PDYTPPEIFSE--STTGFTLYGMLYXPHDLOPKKY 646
 DB 472 ---TKPL-RTITDNKVLQIKNDAVLPNITYFELPLPSGETIANVQRLPPGSPDKKY 525
 QY 647 PVLVFTYGGPOVLVNNRFKGVKK--FRLNTLASIGYVVVINDRGSCHRGILKEGAFKKY 705
 DB 526 PILFTPGGGADEVTWKQALNFKAVVADSELEYVTWVDNRGTGFKRKFRSAVTRQ 585
 QY 706 MGOIIDDQVEGLQYLASRDFIDLDRVGIHGSYGYLSLMALMQRSDIFRVALIAGAPV 765
 DB 586 LGLLEADQIYAAQ--AANIPWIDADHIGIMGMSFGIYLSKYLEKDSGAFTLGVITAPV 644
 QY 766 TLMIFDTGTTEYMGHPDNEQGYIGVYVMAQAEKFPSEPNRLILLHGLIDENVFAFT 825
 DB 645 SDRRFDSMTYTERYKTLSTNEEGYETSAY-RKTDFKQVVEGFTLIQHGTGDNVAFQNS 703
 QY 826 SILSLFVBAKGYDYDIYPOERHSIRVPSGSEHYELH 863

DB 704 AALVDLMDG-----VSPERLHSQWFTDSDHGIGISYH 735
 RESULT 3
 US-10-002-593-6
 ; Sequence 6, Application US/10002593
 ; Patent No. 6586198
 ; GENERAL INFORMATION:
 ; APPLICANT: Vanderbilt University
 ; APPLICANT: Brown, Nancy J
 ; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTING
 ; TITLE OF INVENTION: INHIBITOR AND VASOPEPTIDASE INHIBITOR ASSOCIATED ANGIOEDEMA
 ; FILE REFERENCE: Acty Docker No. 6586198 1242/48/2
 ; CURRENT APPLICATION NUMBER: US/10/002,593
 ; PRIOR FILING DATE: 2001-10-31
 ; PRIOR FILING DATE: 2000-10-31
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 6
 ; LENGTH: 766
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-002-593-6
 Query Match 11.3%; Score 529; DB 4; Length 766;
 Best Local Similarity 26.5%; Pred. No. 2,1e-44;
 Matches 186; Conservative 109; Mismatches 280; Indels 128; Gaps 32;
 QY 206 PADPDM-----AFHSNDIWISNIYREERLLTYVHNELANMEDARSAGVATPV 256
 DB 149 PNNQWYVWSPVGHKLAYVWNNDIYVKLEPILPSYRITWTG-----KEDIITNGITDMV 202
 QY 257 LOEE-PDRYSGYWCRAETTPSGKILRIIYEENDESEVEIITH---VTSMLERRADS 312
 DB 203 YEEVFSAYSLAMSPGTF-----LAVAFNDTEVPLIEVSFSDSLQYKTVR 253
 QY 313 FRYPKGTANPKVTFKXSEIMDAEGRIIDVIDELIOPFEILFEGVYIARAGMTPRG 372
 DB 254 VPYKAGVNPVTAFFV--VNTDSLSSVTNATSQIAPASMLI-GDHYLCDVTA--- 306
 QY 373 YANSILDRSQTLQIVLISPELPIYVEDDVMERORLIESVDSVTPLIYEETDIWIN 432
 DB 307 -----TOERISLOML-----RLQNY--SYMIDICVDESSGRW-N 338
 QY 433 IHDIHFVFPQSHHEIEFIASECKTGFRH-----LYTITSLSKSKYKRSAGLPAP 485
 DB 339 C-----LVAKOHIMSTTGWGRFRPSEPHITLDSNFYKILS--NEEGYRHI----- 384
 QY 486 SDFKCPKEEIAITSGEMEVILGRHSNIOQDEVRLVYFEGT-KDSPLEHLLVYVSVNP 544
 DB 385 CYFOIDKDKCTFITKGTWEVIG---LEALTSYLVYISNEYKMGCGGNLYKIQOLIDY 439
 QY 545 GEYTRLTDROYSHSCCS-CHCPEFISKYSNOKNPH--C---VSLYKLSPEDDPTCK 596
 DB 448 TKVYTCI-----SCEINPERCOYYSVFSKEXKAYYOLRSGGRLPLYTLHSSVNDGLR 492
 QY 597 TKEFWATILDSAGPLPDYTPPE---IFSFTSTGFTLYGMLYXPHDLOPKKYPTVLY 653
 DB 493 VLED-NSALDKM--LQNVQMPSKLDPITINETKF--WQNTLPPH-FDKSKKYPLLLDVY 547
 QY 654 GGPQOVVLNNRFKGVKFRLN--TLASIGYVVV-INDRGSCHRGILKEGAFKKYKQOI 709
 DB 548 AGPCSQ-----KADVFRILNWTYLASTENIIVASPDGSGGYGGDKIHAINRRLGT 601
 QY 710 EIDDOVEGLQYLASRDFIDLDRVGIHGSYGYLSLMALMQRSDIFRVALIAGAPVLT 769
 DB 602 EVADQLEAARQF-SKNGFVNNKRIALWMSYGGYVSMVLGSGSGVFKGCIAAVAPSRME 660
 QY 770 FYDTGYTERYMG--HPDNEQGYIGVYVMAQAEKFPSEPNRLILLHGLIDENVFAFTS 827
 DB 661 YYDSVYTERYMGFLPEDNLDHYRNSVTWSRAENF--KQVEYILLIGTADNVHFOQSAO 718

QY 828 LLSFLVAGKPYDLQIYPOERSIRVPESEGEYELHLLHYLOE 870
 Db 719 ISGALVUVGVDFQAMWTTDEDHGIASTAHQHITHTMSHFITQ 761

RESULT 4

PCT-US93-07923-3

Sequence 3, Application PC/TUS9307923

GENERAL INFORMATION:

APPLICANT: Morimoto, Chikao

APPLICANT: Schlossman, Stuart F.

APPLICANT: Tanaka, Toshiaki

TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB

COMPUTER: IBM PS/2 Model 502 or 55SX

OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

SOFTWARE: Wordperfect (Version 5.0)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/07923

FILING DATE: 19930819

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/934,162

FILING DATE: 21-AUG-1992

APPLICATION NUMBER: 07/832,211

FILING DATE: 06-FEB-1992

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 00530/055002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 755

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

PCT-US93-07923-3

Query Match

Best Local Similarity 26.3%; Score 525; DB 5; Length 755;

Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

206 PADPMI-----AFHSNDIMVSNIVTBEERLTVNHEMLNMEEDASAGVATVY 256

138 PNTQWVWVMSVGHKLAVYMNNDIYKIEPNIPYRITG-----KEDITVNGITDVY 191

257 LOEF-FDRYSGYWCCKAETTPSGGKILRIIYEENDESEVEIIR--VTSPLMLETTRADS 312

192 YEEVEFSAVSALMWSPNGT-----LAVAGNDTEVPLIEVSFYSDESLQYPKYR 242

313 FRYEPTGTANPKVTFRKSEIMIDAGRIIDVDKELIOPELIFGVEXIAPAGVTPBECK 372

243 VVYPAAGVNPVKEFV--VNTDSLSVYVATNSIQTAPASMLI-GDHYLCVTVWA----- 295

373 YAMSILDRSQTRLOIVLISPELFIPEVDVWERQRLIESVDSVTPLIYEFETDIN 432

296 -----TQERISLOWL-----RRIQNY--SVMDICDDESSGRV-N 327

433 IHDIFHVPQSHHEEIEPIFASECKTGRH-----LYKITSIKESKYKSSGGLPAP 485

Db 328 C-----LVANQHEIMSTTGVNVRFPSEPHFTLDGNSFYKIIIS--NEEGYRHI----- 373

QY 486 SDFKCPYKEEIIATSGEWEYLGRHSNIQVDEVRLVYFEGT-KDSPLEHLLVYVSYNP 544

Db 374 CYFOIDKKDCFTFITKGTWEVIG-----TEALTSQVLYIISNEYKMGFGGRNLYKI----- 423

QY 545 GEVTRLTDRGYSHSCCIS-----QHCDFFISKYSNQKPH---C-----VSLYKLSPEDD 592

Db 424 ---QLSD--YTKVYTCISCELPNRCQYVSFVSKEAKYQRLRSRSGGLPLTYTLHSSVND 477

QY 593 PCKTGEFMATIIDSGAPLDIYPPF--IFSESTGFTLYGMLYVPHILOPKKYPTV 649

Db 478 KGRVLED-NSALDKM--LQNVQMPSEKDLFIILNETKF-WYQMIPLPH-FDKSKKPYLL 532

QY 650 LFTYGPQVQVANNRFGVYKFFRLN---TLASLGVVVV--IDNCSGHRGLKFGAPKYY 705

Db 533 LDVYAGPCSQ-----KADTVFRIMATYLASTENITVASFDGSGSYGQKIMHAINR 586

QY 706 MGQIEIDQVGLQYLASRYDFIDLVRGIGHWSYGGYLSLMLMCRSDIFRVAIAGAV 765

Db 587 LGTFEVEDQIEARQF-SKMGFVDNKRIALIMGWSYGGYVSMVLSGSGSYGKGIIVAPV 645

QY 766 TWTIFYDTGYTYRNG--HPDQEGGYLGASVAMQEKFPSEPRNLLHGFLENVHFA 823

Db 646 SRMEYDSVYTERYMGIPPEPDLIDHYRNSTVMSRAENF--KQVEYLLIHGTADNVHFQ 703

QY 824 HTSILSFLVAGKPYDLQIYPOERSIRVPESEGEYELHLLHYLOE 870

Db 704 QSAQISKALVGVDFQAMWTTDEDHGIASTAHQHITHTMSHFITQ 750

RESULT 5

PCT-US93-07923-2

Sequence 2, Application PC/TUS9307923

GENERAL INFORMATION:

APPLICANT: Morimoto, Chikao

APPLICANT: Schlossman, Stuart F.

APPLICANT: Tanaka, Toshiaki

TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB

COMPUTER: IBM PS/2 Model 502 or 55SX

OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

SOFTWARE: Wordperfect (Version 5.0)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/07923

FILING DATE: 19930819

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/934,162

FILING DATE: 21-AUG-1992

APPLICATION NUMBER: 07/832,211

FILING DATE: 06-FEB-1992

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 00530/055002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 759

TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07923-2

Query Match 11.2%; Score 525; DB 5; Length 759;
Best Local Similarity 26.3%; Pred No. 5.4e-44;
Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

QY 206 PADPDM-----AFHSNDIWSNIYREERRLTYVHNLNMEEDARSAGVATFV 256
DB 142 PNTQWMTWSPVGHKLAYVNMNDIYKLEPNLPSRITWTG-----KEDIINYGITWV 195
QY 257 LQEF-FDRYSGYWCPRKAETTPSGKILRILEYENDESEVEIHH---VTSPLMLETREADS 312
DB 196 YEEVFASAYSALWMSPNGT-----LAAQFNDEVPILIEYSFSDLSQYPKTVR 246
QY 313 FRYPKTGTPANKVTFKMSIMIDAGRLLIDVIDKELIOPFELLEGVYIARAGWTPBK 372
DB 247 VPRKAGAVNPYKFFV--VNTDLSSTVNTATSIQITAPASMLI-GDHYLCDVTWA----- 299
QY 373 YAMSLIDRSQTRLQIVLISPELPIVEDDWERORLIESVPDSVTPPLIYEETTDIWIN 432
DB 300 -----TOERISLQWL-----RRIQNV--SVMDICDDESSGRN-N 331
QY 433 IHDFHVPQSHHEEIEFIASECKTGFRH-----LYKITSILKSKYKRSGGCLPAP 485
DB 332 C-----LVARQHLEMTSTGWGRFPSEPHFTLDGNSFYKIIIS--NEGYRHI----- 377
QY 486 SDFKCPKEIEIATSGEWEVLGRHGSNIQVDEVRLLVYFEGT-KDSPLEHHLVYVSYNP 544
DB 378 CYFQIDKKDCFTIKGTWEVIG-----IEALTSDLYIYISNEYKMGNGRNLKYI----- 427
QY 545 GEVTRLDRGYSHSCCIS-----QHCDFFISKYSNOKNPH---C-----VSLYKLSPEDD 592
DB 428 ----QLSD--YTKYTCISCEINPERCQYYSFSKAKYQOLRCSGRLPYLTHSSVND 481
QY 593 PTCKTEFWATILDSAGRLPDYTPPE---IFSFESTTGFTLYGMLYPHDLQPKKPTV 649
DB 482 KGLRVLEB-NSALDKM--LQNVQMPESKLDFTILNETKF--WYQMIILPPH-PDKSKKPYLL 536
QY 650 LFIYGGPOVOLVNNRFKGVKYFRLLN---TLASLGVVVV--IDNRGSGRGLKFGAGAYK 705
DB 537 LDVYAGPCSQ-----KADTVFRLLNMTATYLASTENIIVASFDGSGGQGDKIMHAIINR 590
QY 706 MGQIEIDDOVEGLQYLASRYDFIDLRVGIHGSYGYSLMALMQRSDIFRVAIAGAV 765
DB 591 LGTFEVEDQIEAARQF-SKMGFVDNKRILAIWMSYGYVTSVGLSGSGVFKGIAVAVP 649
QY 766 TLMIFYDTGYTERYMG--HPDQNEGYVIGSVAMQAKFPESENRRLILHGFIDENYHFA 823
DB 650 SRWEYDSYTERYMGFLTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQ 707
QY 824 HTSILSLFLVAKRPYDLQIYPOERHSIRVSEGEHYELHLLHYLOE 870
DB 708 QSAQISKALVDVGDVQAMWYTDHDGIASTAHQHLTYHMSHIKQ 754

RESULT 6

US-08-230-491A-3
Sequence 3, Application US/08230491A
Patent No. 5587299

GENERAL INFORMATION:
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
TITLE OF INVENTION: Gatin-Chesa, Pilar; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED NOCLETIC ACID MOLECULE CODING FOR
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN `AND USES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: FELPE & LYNCH
STREET: 805 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK

COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: IUD 330
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 766 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-230-491A-3

Query Match 11.2%; Score 525; DB 1; Length 766;
Best Local Similarity 26.3%; Pred No. 5.5e-44;
Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

QY 206 PADPDM-----AFHSNDIWSNIYREERRLTYVHNLNMEEDARSAGVATFV 256
DB 149 PNTQWMTWSPVGHKLAYVNMNDIYKLEPNLPSRITWTG-----KEDIINYGITWV 202
QY 257 LQEF-FDRYSGYWCPRKAETTPSGKILRILEYENDESEVEIHH---VTSPLMLETREADS 312
DB 203 YEEVFASAYSALWMSPNGT-----LAAQFNDEVPILIEYSFSDLSQYPKTVR 253
QY 313 FRYPKTGTPANKVTFKMSIMIDAGRLLIDVIDKELIOPFELLEGVYIARAGWTPBK 372
DB 254 VPRKAGAVNPYKFFV--VNTDLSSTVNTATSIQITAPASMLI-GDHYLCDVTWA----- 306
QY 373 YAMSLIDRSQTRLQIVLISPELPIVEDDWERORLIESVPDSVTPPLIYEETTDIWIN 432
DB 307 -----TOERISLQWL-----RRIQNV--SVMDICDDESSGRN-N 338
QY 433 IHDFHVPQSHHEEIEFIASECKTGFRH-----LYKITSILKSKYKRSGGCLPAP 485
DB 339 C-----LVARQHLEMTSTGWGRFPSEPHFTLDGNSFYKIIIS--NEGYRHI----- 384
QY 486 SDFKCPKEIEIATSGEWEVLGRHGSNIQVDEVRLLVYFEGT-KDSPLEHHLVYVSYNP 544
DB 385 CYFQIDKKDCFTIKGTWEVIG-----IEALTSDLYIYISNEYKMGNGRNLKYI----- 434
QY 545 GEVTRLDRGYSHSCCIS-----QHCDFFISKYSNOKNPH---C-----VSLYKLSPEDD 592
DB 435 ----QLSD--YTKYTCISCEINPERCQYYSFSKAKYQOLRCSGRLPYLTHSSVND 488
QY 593 PTCKTEFWATILDSAGRLPDYTPPE---IFSFESTTGFTLYGMLYPHDLQPKKPTV 649
DB 489 KGLRVLEB-NSALDKM--LQNVQMPESKLDFTILNETKF--WYQMIILPPH-PDKSKKPYLL 543
QY 650 LFIYGGPOVOLVNNRFKGVKYFRLLN---TLASLGVVVV--IDNRGSGRGLKFGAGAYK 705
DB 544 LDVYAGPCSQ-----KADTVFRLLNMTATYLASTENIIVASFDGSGGQGDKIMHAIINR 597
QY 706 MGQIEIDDOVEGLQYLASRYDFIDLRVGIHGSYGYSLMALMQRSDIFRVAIAGAV 765
DB 598 LGTFEVEDQIEAARQF-SKMGFVDNKRILAIWMSYGYVTSVGLSGSGVFKGIAVAVP 656
QY 766 TLMIFYDTGYTERYMG--HPDQNEGYVIGSVAMQAKFPESENRRLILHGFIDENYHFA 823
DB 657 SRWEYDSYTERYMGFLTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQ 714

QY 824 HTSILSLFLVRACKPYDLOIYPOERSIRVPESEGEHELHLTYLOE 870
 Db 715 QSAQISKALVDVGVDFOAMWYTDHGHIASTAHQIYTHMSHFIRKQ 761

RESULT 7

US-08-619-280A-3
 ; Sequence 3, Application US/08619280A
 ; Patent No. 5767242
 ; GENERAL INFORMATION:
 ; APPLICANT: Zimmermann, Rainer; Park, John E.;
 ; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
 ; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
 ; TITLE OF INVENTION: ALPHA, AND USES THEREOF
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felte & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/619,280A
 ; FILING DATE: 18-MARCH-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/230,491
 ; FILING DATE: 20-APRIL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hanson, No. 5767242man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 5330.1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 838-3884
 ; TELEFAX: (212) 838-9200
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 766 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-619-280A-3

Query Match 11.2%; Score 525; DB 1; Length 766;

Best Local Similarity 26.3%; Pred. No. 5.5e-44;
 Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33,

QY 206 PADPDMW-----AFHSNDIWSINIVTREERLTYVNEELANNEEDARSAGVATFV 256
 Db 149 PNNQWVWSPVGHKLAVYNNNDIYVKLEPNLPSYRITWTG-----KEDIYNGIDMW 202
 QY 257 LQEE-FDRYSGYWGCKKATTPSSGKILRIIYEENDESEVNIH---VTSMLETRADS 312
 Db 203 YEEVFSAISALWSPNGTF-----LVAQPNDEVLPLEISFYSDSLSLOYPKTVR 253
 QY 313 FRYPKTANPKVTPKMSIMIDAGRIIDVIDKLIOPFELFEGVEYIARAQMTDEGK 372
 Db 254 VYPRAGAVNPTVKFV--VNTDLSSTVNTATSIQITAPAMLI-GDHYLCDVYWA---- 306
 QY 373 YAWSILLDRSQTRLOIYLISBELFLPVEDDVMERQRLIESVPDSVTPLIYEETDIWIN 432
 Db 307 -----TQERISLQWL-----RRLQNY--SYMDICDDESSGRW-N 338
 QY 433 IHDIHFVPSGHEEIEIFASECTGFRH-----LYKTSILKESKYRSGGAP 485
 Db 339 C-----LVARQHIEMSTGTGWVGRFRSPHFTLDGNSFYKIIIS--NEGYNHIL----- 384
 QY 486 SDFKPIKEBIAITSGEWEVLGRHSNIQVDEVRRLVYFEGT-KDSPLEHLLVYVSYNP 544

Db 385 CYFQIDKKDCTPITTKTWEVIG-----TEALSDIYIISNEYKMPGGRNLTKI----- 434
 QY 545 GEVTRLTDRGYSKSCIS-----QHCFPIRSKYNQKNP---C-----VSLYKLSPEDD 592
 Db 435 -----QLSD--YTKVTCISCELNERCQYYSVFSFKAKYQJLRCSGPGJPLTYLHSSVND 488
 QY 593 PTCCKEPMATILDSAGPLPDYTPRE---IPSESTGTGLYMLKPHDLQPKKYPV 649
 Db 489 KGLRVLED-NSALDKM--LQNVQMPSKIDPFIILNETKF-WYQMLPDP-PDKSKYPLL 543
 QY 650 LPIYGPQVQVLVNNRFGYKFFLN---TLASIGYVVV-IDNRGSHRGLKEGAFKKY 705
 Db 544 LDVYAGPCQ-----KADTVRELMATYLAENENIVASFGRSGCYGDKIMAINR 597
 QY 706 MGQIEIDQVEGLQYIASRYDPIIDLVRVIGHWSYGGYLSLALMORSDFRVALIAGPV 765
 Db 598 LGTFEEVDQTEARQF-SKQGFVDNRIRIAIWGSYGGYVTSWLGSGSGVFKCGIYAVPV 656
 QY 766 TLMIFYDQTYTERYMG--HPDQNEGGYILGSYVAMQAEKFPSEBNRLLLHGFLENVHA 823
 Db 657 SRWEYDSYTERYMGELPTPEDNLDHYRSTVMSRAENF--KQVEYLLHGTADNVHQQ 714
 QY 824 HTSILSLFLVRACKPYDLOIYPOERSIRVPESEGEHELHLTYLOE 870
 Db 715 QSAQISKALVDVGVDFOAMWYTDHGHIASTAHQIYTHMSHFIRKQ 761

RESULT 8

US-08-940-391-3
 ; Sequence 3, Application US/08940391
 ; Patent No. 5965373
 ; GENERAL INFORMATION:
 ; APPLICANT: Zimmermann, Rainer; Park, John E.;
 ; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
 ; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
 ; TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felte & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/940,391
 ; FILING DATE: 01-OCT-1997
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/619,280
 ; FILING DATE: 18-MARCH-1996
 ; APPLICATION NUMBER: 08/230,491
 ; FILING DATE: 20-APRIL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hanson, No. 5965373man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 5330.1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 838-3884
 ; TELEFAX: (212) 838-9200
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 766 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-940-391-3

Query Match 11.2%; Score 525; DB 2; Length 766;
 Best Local Similarity 26.3%; Pred. No. 5, 5e-44;
 Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

```

QY 206 PADPDMT-----AFHSNDIWSNVTREERLTGVHNLAMMEDASAGVATPV 256
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 PNNQWVTVSPVGHKLAYWNNDIYKIEPNLPSYRITWTG-----KEDIYNGITD 202
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 LQEE-PDRYSGWVCKPKEATTSGGKILRLYEENDESEVEIHH--VTSPLMLET 312
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 YEEVFSAYSALWMSPNGT-----LAVAQFNDEVPLEIYFSYDESLOVYKTV 253
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 313 FRYPKGTANPKVTEKMEIMIDAEGRIIDVDKELIOPFELIFEGVEYIARAGMTPE 372
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 VYPRKAGVNPVYKFFV--VNTDSLSVTNATSIQTAPASMLI--GDHYLCVYTA 306
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 373 YAMSLIDRSQTRLQIVLISPELIFVEDDVMERQRLIESVDSTVPLTIYEETDI 432
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 -----TORISLOWL-----RRIQNY--SVWDICDYESSGRW-N 338
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 433 IHDIHVPPQSHHEEIEIFASECKTGRH-----LYKTSILKESYKRSGGLPAP 485
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 C-----LVARQHIEMSTTGMVGRFPSPHFTLDGNSFYKIIIS--NEEGYRHI 384
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 SDFKCPIKEEIAITSGEMEVLRHGSNIQVDEVRRLVFEET-KDSPLEHLLYVVS 544
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 385 CYFOIDKDCDCTFITGTWEVIG-----IEALTSVLYIISNEYKMPGGRNLKYI 434
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 545 GEVTRLTRGYSHSCCIS-----QHDFEISKYSNQKNPH--C-----VSLYK 592
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 ----QLSD--YTKVYCLSCELNPERCQYVSFSKAKYQLRCSGPGPLPYTLH 488
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 593 PTCKTEFWATLIDAGPLPDYTPPE---ISFESTGTFTLYGMKYKHDLPGKKYPT 649
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 489 KGLRPLED-NSALDQM--LQNVQMPSKKLDPIILNETKF-WYQMLPRH-PDK 543
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 650 LFIYGGPOVOLVNNRFGKVKYFRLN---TLASLGVVVV-IDNRGSGHGLKFE 705
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 544 LDVYAGPCSQ-----KADYVRLNMAITYLASTENIIVASFDGRSGYQGD 597
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 706 MGOEIIDQVQGLQYLASRYDFIDLRVGIHGSYGYLSMALMQRSDFRVAIAGAP 765
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 598 LGTEFEVDQIEAARQF-SKMGFVDNKRIAIWGMSTYGVYTMVIGSSGVFC 656
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 766 TLMIFYDTGYTERYNG--HPDQNEGYVLSVANMAEKFPSEPNRLILHGFLEN 823
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 657 SRMEYISVYTERIMGLPTPEDNDHYRNSTWMSRAENF--KOVEYLLIHGTAD 714
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 824 HTSILSFLVRAKRPYDLQIYPOERHSIRVESGEHYELHLAYLOE 870
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 715 QSAQISKALVDVGVDFQAMWYTTDEDHGIASSTAHQHITYHMSHPKIQ 761
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9
 US-09-794-236-1
 ; Sequence 1, Application US/09794236
 ; Patent No. 6337069
 ; GENERAL INFORMATION:

; APPLICANT: Grouzmann, Eric
 ; APPLICANT: Lacroix, Jean-Silvain
 ; APPLICANT: Monod, Michel
 ; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
 ; FILE REFERENCE: 81985/276823
 ; CURRENT APPLICATION NUMBER: US/09/794.236
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: Patentin version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 766
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 ; US-09-794-236-1

Query Match 11.2%; Score 525; DB 4; Length 766;
 Best Local Similarity 26.3%; Pred. No. 5, 5e-44;
 Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

```

QY 206 PADPDMT-----AFHSNDIWSNVTREERLTGVHNLAMMEDASAGVATPV 256
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 PNNQWVTVSPVGHKLAYWNNDIYKIEPNLPSYRITWTG-----KEDIYNGITD 202
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 LQEE-PDRYSGWVCKPKEATTSGGKILRLYEENDESEVEIHH--VTSPLMLET 312
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 YEEVFSAYSALWMSPNGT-----LAVAQFNDEVPLEIYFSYDESLOVYKTV 253
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 313 FRYPKGTANPKVTEKMEIMIDAEGRIIDVDKELIOPFELIFEGVEYIARAGMTPE 372
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 VYPRKAGVNPVYKFFV--VNTDSLSVTNATSIQTAPASMLI--GDHYLCVYTA 306
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 373 YAMSLIDRSQTRLQIVLISPELIFVEDDVMERQRLIESVDSTVPLTIYEETDI 432
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 -----TORISLOWL-----RRIQNY--SVWDICDYESSGRW-N 338
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 433 IHDIHVPPQSHHEEIEIFASECKTGRH-----LYKTSILKESYKRSGGLPAP 485
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 C-----LVARQHIEMSTTGMVGRFPSPHFTLDGNSFYKIIIS--NEEGYRHI 384
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 SDFKCPIKEEIAITSGEMEVLRHGSNIQVDEVRRLVFEET-KDSPLEHLLYVVS 544
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 385 CYFOIDKDCDCTFITGTWEVIG-----IEALTSVLYIISNEYKMPGGRNLKYI 434
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 545 GEVTRLTRGYSHSCCIS-----QHDFEISKYSNQKNPH--C-----VSLYK 592
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 ----QLSD--YTKVYCLSCELNPERCQYVSFSKAKYQLRCSGPGPLPYTLH 488
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 593 PTCKTEFWATLIDAGPLPDYTPPE---ISFESTGTFTLYGMKYKHDLPGKKYPT 649
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 489 KGLRPLED-NSALDQM--LQNVQMPSKKLDPIILNETKF-WYQMLPRH-PDK 543
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 650 LFIYGGPOVOLVNNRFGKVKYFRLN---TLASLGVVVV-IDNRGSGHGLKFE 705
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 544 LDVYAGPCSQ-----KADYVRLNMAITYLASTENIIVASFDGRSGYQGD 597
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 706 MGOEIIDQVQGLQYLASRYDFIDLRVGIHGSYGYLSMALMQRSDFRVAIAGAP 765
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 598 LGTEFEVDQIEAARQF-SKMGFVDNKRIAIWGMSTYGVYTMVIGSSGVFC 656
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 766 TLMIFYDTGYTERYNG--HPDQNEGYVLSVANMAEKFPSEPNRLILHGFLEN 823
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 657 SRMEYISVYTERIMGLPTPEDNDHYRNSTWMSRAENF--KOVEYLLIHGTAD 714
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 824 HTSILSFLVRAKRPYDLQIYPOERHSIRVESGEHYELHLAYLOE 870
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 715 QSAQISKALVDVGVDFQAMWYTTDEDHGIASSTAHQHITYHMSHPKIQ 761
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10
 US-08-230-491A-2
 ; Sequence 2, Application US/08230491A
 ; Patent No. 5587299
 ; GENERAL INFORMATION:

; APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
 ; APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
 ; TITLE OF INVENTION: ISOLATED NOCLEIC ACID MOLECULE CODING FOR
 ; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN ` AND USES
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: FELFE & LYNCH
 ; STREET: 805 THIRD AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE

```

1 COMPUTER: IBM PS/2
2 OPERATING SYSTEM: PC-DOS
3 SOFTWARE: WORDPERFECT - ASC II
4 CURRENT APPLICATION DATA:
5 APPLICATION NUMBER: US/08/230,491A
6 FILING DATE: 20-APRIL-1994
7 CLASSIFICATION: 435
8 ATTORNEY/AGENT INFORMATION:
9 NAME: Hanson, No. 5587299man D.
10 REGISTRATION NUMBER: 30,946
11 REFERENCE/DOCKET NUMBER: LUD 330
12 TELECOMMUNICATION INFORMATION:
13 TELEPHONE: (212) 688-9200
14 TELEFAX: (212) 838-3884
15 INFORMATION FOR SEQ ID NO: 2:
16 SEQUENCE CHARACTERISTICS:
17 LENGTH: 760 amino acids
18 TYPE: amino acid
19 TOPOLOGY: linear
20
21 US-08-230-491A-2

```

```

Db 431 SYPSKCCVCHLKRERCOYTTASDYAKYALVCYGPPISTLHDGRDQE-----IK 486
Qy 604 ILDSAGLPD-----YTPPEIFSESTTGFTLYGMLYKPHDLOPKKYPVLYFYGGPQV 658
Db 487 ILEENKELEMLAKNIQLPKEIKLEVEDEITLWYKMIIPPOFDRSKKYPPLIQYGGPCS 546
Qy 659 QLVNRRFGVKYFRINTLASL---GYVVVVIDNRGSGHGLKEFGAFKYMGOIEIDDQ 714
Db 547 QSVRS-----VFAVNMISTYLSKSGGVIALVDGRGTAFQGDKLLVAVYRLGYVEVEDQ 600
Qy 715 VEGLOYLASRYDFIDLDVGIHGMVSGYLSLMLMQSDIFRVAIAGAPVTLTFYDTG 774
Db 601 ITAVRKFI-EMGFIDEKRIALWMSYGYVSSIALASGTGLFKGCIATAVAPSSMEYASV 659
Qy 775 YTERYMGHP--DQNEQGYVIGSVAMQAEKPESEPNRLLHGFIDENVHFAHTSILSFL 832
Db 660 YTERFMGLPTKDNLEHKNSTVWARAEYFRNVD--YLLHGTADNVHFONSAQIAKAL 717
Qy 833 VRAGKPYDLQIYPOERHSIRVPESG---EYELHLHYLOE 870
Db 718 VNAQVDFQAMWYSDONHGL---SGLSTNHLTYHMTHTFLKQ 754

```

RESULT 12

```

US-08-940-391-2
; Sequence 2, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainert; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.;
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; NUMBER OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,391
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/619,280
; FILING DATE: 18-MARCH-1996
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5965373man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-940-391-2

```

```

Query Match          9.7%; Score 454; DB 2; Length 760;
Best Local Similarity 24.1%; Freq. No. 9.5e-37;
Matches 169; Conservative 113; Mismatches 269; Indels 150; Gaps 27;

```

```

Qy 212 IAFHSNDIWNISNIVTREERLTYVHNELANMEEDASAGVATFVLOEFD--DRYSGYWM 269
Db 162 LAYVQNNIYIKQRPGRPPFOITP-----NGRENKIPNGIPDMVYEEMLPTKYA-LWM 214
Qy 270 CPKAEPTPSGGKILRIIYEENDESEVELIHVTSPLMET-RRASFRYPKGTANPKVTFK 328
Db 215 SP-----NGKFL--AAEFNDKIDIPVIAVSYGDEQYPTINIPYKAKAKNPVARI- 264
Qy 329 MSEIMDAEGRIIVIDIKELIQPEI-----LFGVEYIAPAGVTPGKYMMSILDRSQ 383
Db 265 -----FIDITTPAYVGPQEVVPVAMIASDSYFVSWLTWTDERVCLQWL----- 309
Qy 384 TRLOIVLISPLRFIVEDDQWVERQRLTESVDSVT-----PLIYETTDIWNINI 433
Db 310 KRVQNVSVLSICDRPDQWTDCKTOEHIEESTIGWAGFVSRPVSYSYDA-----ISY 364
Qy 434 HDIHFVPPQSHHEEIEFIASECKTGRHLKYITSILKESKYSKSSGGLPAPSPFKCPK 493
Db 365 YKIF-----SDKDYKHIHYI-----KDTIVE 385
Qy 494 BEIATSGEWEVLGRHSNIOVDEVRRLVYFEGTKDSPLEHNLVYVSIVNPGEVTRLTDR 553
Db 386 NAIQITSGKWEAL-----NT-----FRVQDSLFSYSSNEFEYPGRRNRYRISIG 430
Qy 554 GYSHS-CCISQH-----CDFEISKYSNQKAPHCVSLY-----KLSPPDDPTCKTKEFAT 603
Db 431 SYPSKCCVCHLKRERCOYTTASFSYAKYALVCYGPPISTLHDGRDQE-----IK 486
Qy 604 ILDSAGLPD-----YTPPEIFSESTTGFTLYGMLYKPHDLOPKKYPVLYFYGGPQV 658
Db 487 ILEENKELEMLAKNIQLPKEIKLEVEDEITLWYKMIIPPOFDRSKKYPPLIQYGGPCS 546
Qy 659 QLVNRRFGVKYFRINTLASL---GYVVVVIDNRGSGHGLKEFGAFKYMGOIEIDDQ 714
Db 547 QSVRS-----VFAVNMISTYLSKSGGVIALVDGRGTAFQGDKLLVAVYRLGYVEVEDQ 600
Qy 715 VEGLOYLASRYDFIDLDVGIHGMVSGYLSLMLMQSDIFRVAIAGAPVTLTFYDTG 774
Db 601 ITAVRKFI-EMGFIDEKRIALWMSYGYVSSIALASGTGLFKGCIATAVAPSSMEYASV 659
Qy 775 YTERYMGHP--DQNEQGYVIGSVAMQAEKPESEPNRLLHGFIDENVHFAHTSILSFL 832
Db 660 YTERFMGLPTKDNLEHKNSTVWARAEYFRNVD--YLLHGTADNVHFONSAQIAKAL 717
Qy 833 VRAGKPYDLQIYPOERHSIRVPESG---EYELHLHYLOE 870
Db 718 VNAQVDFQAMWYSDONHGL---SGLSTNHLTYHMTHTFLKQ 754

```

RESULT 13

```

US-09-016-080-1
; Sequence 1, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Matsui, Ikuro
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshisugu
; APPLICANT: Higuchi, Katsuhiko
; TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; EARLIER FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 1
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-09-016-080-1

```

Query Match 6.0%; Score 282; DB 3; Length 632;

Best Local Similarity 23.9%; Pred. No. 2.4e-19;

Matches 151; Conservative 91; Mismatches 242; Indels 148; Gaps 33;

276 TPGSGKILRIEYENDESEVEIHHVTSPLMTERADSPRYPKTGTANPKYTFKMSIM- 333

67 SPNGLIAPTSKRDEKESSELYVPTDGEARLLAKFY---GKILRPFEDOKSLAV 123

334 --IDAEGRIIDVIDKELLQPELIFEGVEYIARAGWTEGKYAMSLD--RSQYTL- 386

124 TPIDVEKKGN--DVHIIREIIPFWNGV---GWI-YGRKNVYLVDSGCKKRLTP 174

387 -----QVILSPLEIFIVEDD-----VMERQ-----RLIESVPDVT 419

175 KNLAVDQRPFHNGRIYFAQEDREKPLISDLYLENKVKRLTPGKRLIDPLFDGGS 234

420 LIIEETD--IWINIHDFHVPQSHHEIEIIFASECKTGFHLYKITSILKSKYK 477

235 FVLKANTLERGIPTNAH-IYHYDP-----KTG--ELKKLTGYDLDRMAYNS 276

478 SSGGLPAPSDPKCPKEKRIATSGEMVEYGRHSNIOVDEVRALVFEQTKDSPLEH 537

277 LN-----SDVRGSGRAELVYKEG-W-----IYVAT-DGP-RANLP 309

538 VSYVNPGEVTRLT--DRGYSHSCISQHCDFISKYSNQNPKHCVSLYKLSPEDDPTC 595

310 RVNL--DGKIERVIGGDSVE-----SFDIGDYIAFAQAVPTPELTLYRDKER 358

596 KTKER--WATILDSAGRLPDYT--PRLIFSESTTGTLGYMLYKPHLDQPKKYPVLF 651

359 KVTDPNKV-----IKGYTLKPEHFKVADGVEIDAMVWKPVNRFGKKYPAILE 409

652 IYGGPOVOLVNNREFKVKYF--RLNTLASLGYVVVVINDRSGCHGLKFEAGFKYK 710

410 IHGKFKYV-----GYAFMEHFVILSKGFVVFISNRSDDGGEF--ADIRGHYERD 462

711 IDDOVEGLQYLAASYDFTLDRVGIHWSYGYISLMAIMORSDFRVALAGAPVTLMF 770

463 YQDLMEVVDALRRDFIDGERLGYTGSGYGFMTNW--IVGHTNRFKAATVQSRISNMIS 521

771 Y-----DNGYTERYMGHP-----NEOGYILGSVAMQAEKPESEPNRILLHGFIDEN 819

522 FFGTTDGL--YFA-PDQIGKDPKSNLEGYWEKSPKYA--DNVETPLILHSTEDYR 574

820 VHPAFTSILSFLVRAKRPYDLOIYQPERHSI 851

575 CWLPEALQFLSLKYLGRVLELFPGENHDL 606

RESULT 14

US-09-355-166-1

; Sequence 1, Application US/09355166

; Patent No. 6316241

; GENERAL INFORMATION:

; APPLICANT: Genencor International, Inc.

; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes

; FILE REFERENCE: GC511-PCT

; CURRENT APPLICATION NUMBER: US/09/355,166

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 657

; TYPE: PRT

; ORGANISM: Bacillus

Query Match 5.3%; Score 249.5; DB 4; Length 657;

Best Local Similarity 21.4%; Pred. No. 5.3e-16;

Matches 150; Conservative 101; Mismatches 265; Indels 185; Gaps 31;

245 EDASAGVATVVLQEEFDRSGYMWCPKAEY-----TPSGKILRI 286

DB 24 DGTAAVYKSVQVQEK-DSYTSNIMVYETKGTGSPWTHGKRSSTDPRKSPGRLTAS 82

QY 287 YEENDESEVEIHHVTSPLMTERADSPRYPKTGTANPKYTFKMSIMID--AEGRIIDV 343

DB 83 DREGDAQAQYIMSTEGG--EARKLTDIPY---GVSKPLMSPGESILVTISLGESESID 137

QY 344 IDK---ELIQPELIFEGVEYIARAGWTEGKYAMSLD--RSQYTLQVILSPLEI 400

DB 138 REKTEQDSYEPVEV--QGLSYRDKGLTRGVA-----QVLVSVK----- 177

QY 401 DVMERQRLIESVPDVTPL-----IIEET--TDIWINIHDFHVPQSHHEIEF 450

DB 178 --SGEMKELTSKADHGDPAFSPDKWLVFSANLEITDASKPHVY-----LMS 225

QY 451 IPASECK--TGRPHLYKITSILKSKYKSSGGLPAPSDPKCPKEKRIATSGEM- 503

DB 226 LESGDKQYTPHRSFGSSSFSPDKRYALLGN--EKEYK-----NATLSKAMLYDIE 276

QY 504 -----EVLGRH-----GSNIQDEVRLVFEQTKDSPLEHLYVVSYNPGEVTR 549

DB 277 QGRULTLTMVHLADALIGDSLIGAEQRIW--TKDS--OGFYVIG----- 321

QY 550 LTRDG-----YSHSCISQHCDFISKYSNQNPKHCVSLYKLS 588

DB 322 --TDQSGTGYIYISBGLVYPIREKEKYNISFELSPDEQGFISYTKPRPSELISIPLOQ 380

QY 589 PEDDPCKRKEPMATILDSAGRLPDYT--PRLIFSESTTGTLGYMLYKPHLDQPKKY 646

DB 381 BEKQLTGANDKF-----VREHTTISPEIQAATEDGVAMGMLRPPQMGERTY 430

QY 647 PTVLPYGGPOVOLVNNREFKVKYF--RLNTLASLGYVVVVINDRSGCHGLKFEAGFKYK 705

DB 431 PLINIHGPHMW-----GHTYFHEPQVLAKGYAVVYINPRSHGQGFVANAVGD 484

QY 706 MGQIETDDQVEGLQYLAASYDFTLDRVGIHWSYGYISLMAIMORSDFRVALAGAPV 765

DB 485 YGKGDVDMQAVDEAIKDPHIDPKRLGVTGSGYGFMTNW--IVGHTNRFKAATVQSRIS 543

QY 766 TLMIFY-----DNGY--TERYMGHP-----DNEOGYILGSVAMQAEKPESEPNRLL 810

DB 544 SNWISHGVSIDGYPTDQLHDMFEDTEKLMDSPLKY-----AANYE-----TPLL 592

QY 811 LHHGFIDENVHFAHTSILSFLVRAKRPYDLOIYQPERHSI 851

DB 593 ILHGERDRCPIBOAEOLFIALKKMKGETKLVFPNASHNL 633

RESULT 15

PCT-US93-07923-11

; Sequence 11, Application PC/TUS9307923

; GENERAL INFORMATION:

; APPLICANT: Morimoto, Chikao

; APPLICANT: Schlossman, Stuart F.

; APPLICANT: Tanaka, Toshiaki

; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

ZIP: 02110-8804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB

COMPUTER: IBM PS/2 Model 502 or 555X

OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

SOFTWARE: WordPerfect (Version 5.0)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/07923

FILING DATE: 19930819

CLASSIFICATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:13:57 ; Search time 124.626 Seconds
(without alignments)
1140.337 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700
Sequence: 1 MAAMETEQLGVEIFETADC.....HLHYIQENIGSRNALKVI 882

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4700	100.0	882	10	US-09-976-674-1 Sequence 1, Appl1
2	4700	100.0	882	12	US-10-054-776-2 Sequence 2, Appl1
3	4700	100.0	882	12	US-10-170-789-38 Sequence 38, Appl1
4	3504	74.6	658	10	US-09-976-674-19 Sequence 19, Appl1
5	3504	74.6	661	10	US-09-976-674-11 Sequence 11, Appl1
6	3504	74.6	690	10	US-09-976-674-71 Sequence 71, Appl1
7	3236	66.9	613	10	US-09-976-674-21 Sequence 21, Appl1
8	2870	61.1	863	10	US-09-976-674-3 Sequence 3, Appl1
9	2870	61.1	892	10	US-09-976-674-23 Sequence 23, Appl1
10	2870	61.1	892	10	US-09-976-674-27 Sequence 27, Appl1
11	2820.5	60.0	879	10	US-09-976-674-35 Sequence 35, Appl1
12	2820.5	60.0	879	10	US-09-976-674-33 Sequence 33, Appl1
13	2406	51.2	832	10	US-09-976-674-29 Sequence 29, Appl1
14	2406	51.2	832	10	US-09-976-674-31 Sequence 31, Appl1
15	2356.5	50.1	819	10	US-09-976-674-37 Sequence 37, Appl1

16	2356.5	50.1	819	10	US-09-976-674-39 Sequence 39, Appl1
17	1808	38.5	358	10	US-09-976-674-13 Sequence 13, Appl1
18	1645.5	35.0	310	12	US-09-993-959-4 Sequence 4, Appl1
19	1643.5	35.0	508	15	US-10-237-271-3 Sequence 3, Appl1
20	1391	29.6	518	10	US-09-976-674-25 Sequence 25, Appl1
21	1278	27.2	241	10	US-09-976-674-9 Sequence 9, Appl1
22	1049	22.3	432	15	US-10-237-271-4 Sequence 4, Appl1
23	1007.5	21.4	194	10	US-09-976-674-17 Sequence 17, Appl1
24	585	12.4	129	11	US-09-764-891-3564 Sequence 3564, Ap
25	529	11.3	766	12	US-10-423-714-6 Sequence 6, Appl1
26	529	11.3	766	14	US-10-007-593-6 Sequence 6, Appl1
27	529	11.3	766	15	US-10-165-603-7 Sequence 7, Appl1
28	525	11.2	766	9	US-09-265-608-3 Sequence 3, Appl1
29	525	11.2	766	12	US-09-993-959-1 Sequence 1, Appl1
30	523.5	11.1	767	15	US-10-165-603-4 Sequence 4, Appl1
31	495	10.5	710	15	US-10-156-761-10681 Sequence 10681, A
32	484.5	10.3	818	12	US-10-401-437-3 Sequence 3, Appl1
33	484.5	10.3	818	12	US-10-402-067-3 Sequence 3, Appl1
34	484.5	10.3	818	12	US-10-401-436-3 Sequence 3, Appl1
35	484.5	10.3	818	12	US-09-976-674-15 Sequence 15, Appl1
36	464	9.9	108	10	US-09-976-674-41 Sequence 41, Appl1
37	461	9.8	706	10	US-09-976-674-5 Sequence 5, Appl1
38	461	9.8	796	10	US-09-976-674-5 Sequence 5, Appl1
39	461	9.7	760	12	US-10-301-822-55 Sequence 55, Appl1
42	454	9.7	760	15	US-10-177-253-136 Sequence 136, App
43	454	9.7	803	12	US-10-205-219-58 Sequence 58, Appl1
44	431	9.2	691	10	US-09-976-674-43 Sequence 43, Appl1
45	425.5	9.1			

ALIGNMENTS

```
RESULT 1
US-09-976-674-1
; Sequence 1, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OR INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-1
```

Query Match	100.0%	Score 4700;	DB 10;	Length 882;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 882;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSNQLKLLADTRKYGVNM	60	
DB	1	MAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSNQLKLLADTRKYGVNM	60	
QY	61	AKAPHFMFVKRNDPPGPHSDRIYIAMSNGENRENTLFFSEIPIKTIINRAAVMLSKPIL	120	
DB	61	AKAPHFMFVKRNDPPGPHSDRIYIAMSNGENRENTLFFSEIPIKTIINRAAVMLSKPIL	120	
QY	121	DLFQATLDYGMYSREBELRRERRIGTVGIAVDYHQSGTFLFQAGSIYHVKGQGPQ	180	
DB	121	DLFQATLDYGMYSREBELRRERRIGTVGIAVDYHQSGTFLFQAGSIYHVKGQGPQ	180	

```

QY      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIIVTREERLLTYVHNEI 240
Db      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIIVTREERLLTYVHNEI 240
QY      241 ANMEEDASAGVATFVLOEFPDRYSGYWCCKAETTPSGKILRLIYEENDESEVEIIVH 300
Db      241 ANMEEDASAGVATFVLOEFPDRYSGYWCCKAETTPSGKILRLIYEENDESEVEIIVH 300
QY      301 TSPMLSTRADSFRRPKGTANPKYTFKMSIIMDAEGRIIDVIDKELIQPELIFEGVE 360
Db      301 TSPMLSTRADSFRRPKGTANPKYTFKMSIIMDAEGRIIDVIDKELIQPELIFEGVE 360
QY      361 YIARAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERORLIESVDSVPL 420
Db      361 YIARAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERORLIESVDSVPL 420
QY      421 ILYEETTDIMINIHDI FHVFPQSHHEEIEFI FASECKTGFRHLKYITSLKESKYKSSG 480
Db      421 ILYEETTDIMINIHDI FHVFPQSHHEEIEFI FASECKTGFRHLKYITSLKESKYKSSG 480
QY      481 GLPAPSDFKCPKIEKIALITSGEMVYLGHRGNSIQVDEVRLVYFEGTKDSPLEHLLYVS 540
Db      481 GLPAPSDFKCPKIEKIALITSGEMVYLGHRGNSIQVDEVRLVYFEGTKDSPLEHLLYVS 540
QY      541 YVNPGEVTRLTDRGYSHSCCISOHCDPFIISKYSNQKNPCHVSLYKLSPEDDPTCKTKEF 600
Db      541 YVNPGEVTRLTDRGYSHSCCISOHCDPFIISKYSNQKNPCHVSLYKLSPEDDPTCKTKEF 600
QY      601 WATILDSAGPLPDYTPPELIFSFESTTGFTLYGMLYKPHDLOPKKYPTVLFIYGGPOVOL 660
Db      601 WATILDSAGPLPDYTPPELIFSFESTTGFTLYGMLYKPHDLOPKKYPTVLFIYGGPOVOL 660
QY      661 VNNRFGVYKFRPLNTLASLGYVVVVYIDNRGSCHRGLKEGAFKXKMGQIEIDDOVEGLQY 720
Db      661 VNNRFGVYKFRPLNTLASLGYVVVVYIDNRGSCHRGLKEGAFKXKMGQIEIDDOVEGLQY 720
QY      721 LASRYDFIDLRVGIHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFDTGYTERYM 780
Db      721 LASRYDFIDLRVGIHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFDTGYTERYM 780
QY      781 GHPDNEQGYLYGSVAMQAEKFPSEPNRLLLHGFIDENVAFAHTSILSLVRAKPYD 840
Db      781 GHPDNEQGYLYGSVAMQAEKFPSEPNRLLLHGFIDENVAFAHTSILSLVRAKPYD 840
QY      841 LQIYPOERHSIRVPESGHEYLHLHYIQENIGSRIALKVI 882
Db      841 LQIYPOERHSIRVPESGHEYLHLHYIQENIGSRIALKVI 882

```

RESULT 2

```

US-10-054-776-2
; Sequence 2, Application US/10054776
; Publication No. US20030165818A1
; GENERAL INFORMATION:
; APPLICANT: Mark Robert Edbrooke
; APPLICANT: Alan Peter Lewis
; TITLE OF INVENTION: NOVEL PROTEIN
; FILE REFERENCE: QG10420S
; CURRENT APPLICATION NUMBER: US/10/054.776
; CURRENT FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-054-776-2

```

```

Query Match      100.0%; Score 4700; DB 12; Length 882;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 MAAMETDQGYEIPETADCEENISODRPKLEPFYVERYSQIKLLADTRKHGYMM 60
Db      1 MAAMETDQGYEIPETADCEENISODRPKLEPFYVERYSQIKLLADTRKHGYMM 60
QY      61 AKAPDFMFKVKNDDPGPHSDRIYLLAMSGENRENTLIFYSEIPIKTIINBAALVLMISKPL 120
Db      61 AKAPDFMFKVKNDDPGPHSDRIYLLAMSGENRENTLIFYSEIPIKTIINBAALVLMISKPL 120
QY      121 DLFOATLDYGMYSREBELIREKRIGTVGIASVDYHQSGLFLFOAGSGIYHVKGQFQG 180
Db      121 DLFOATLDYGMYSREBELIREKRIGTVGIASVDYHQSGLFLFOAGSGIYHVKGQFQG 180
QY      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIIVTREERLLTYVHNEI 240
Db      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIIVTREERLLTYVHNEI 240
QY      241 ANMEEDASAGVATFVLOEFPDRYSGYWCCKAETTPSGKILRLIYEENDESEVEIIVH 300
Db      241 ANMEEDASAGVATFVLOEFPDRYSGYWCCKAETTPSGKILRLIYEENDESEVEIIVH 300
QY      301 TSPMLSTRADSFRRPKGTANPKYTFKMSIIMDAEGRIIDVIDKELIQPELIFEGVE 360
Db      301 TSPMLSTRADSFRRPKGTANPKYTFKMSIIMDAEGRIIDVIDKELIQPELIFEGVE 360
QY      361 YIARAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERORLIESVDSVPL 420
Db      361 YIARAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERORLIESVDSVPL 420
QY      421 ILYEETTDIMINIHDI FHVFPQSHHEEIEFI FASECKTGFRHLKYITSLKESKYKSSG 480
Db      421 ILYEETTDIMINIHDI FHVFPQSHHEEIEFI FASECKTGFRHLKYITSLKESKYKSSG 480
QY      481 GLPAPSDFKCPKIEKIALITSGEMVYLGHRGNSIQVDEVRLVYFEGTKDSPLEHLLYVS 540
Db      481 GLPAPSDFKCPKIEKIALITSGEMVYLGHRGNSIQVDEVRLVYFEGTKDSPLEHLLYVS 540
QY      541 YVNPGEVTRLTDRGYSHSCCISOHCDPFIISKYSNQKNPCHVSLYKLSPEDDPTCKTKEF 600
Db      541 YVNPGEVTRLTDRGYSHSCCISOHCDPFIISKYSNQKNPCHVSLYKLSPEDDPTCKTKEF 600
QY      601 WATILDSAGPLPDYTPPELIFSFESTTGFTLYGMLYKPHDLOPKKYPTVLFIYGGPOVOL 660
Db      601 WATILDSAGPLPDYTPPELIFSFESTTGFTLYGMLYKPHDLOPKKYPTVLFIYGGPOVOL 660
QY      661 VNNRFGVYKFRPLNTLASLGYVVVVYIDNRGSCHRGLKEGAFKXKMGQIEIDDOVEGLQY 720
Db      661 VNNRFGVYKFRPLNTLASLGYVVVVYIDNRGSCHRGLKEGAFKXKMGQIEIDDOVEGLQY 720
QY      721 LASRYDFIDLRVGIHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFDTGYTERYM 780
Db      721 LASRYDFIDLRVGIHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFDTGYTERYM 780
QY      781 GHPDNEQGYLYGSVAMQAEKFPSEPNRLLLHGFIDENVAFAHTSILSLVRAKPYD 840
Db      781 GHPDNEQGYLYGSVAMQAEKFPSEPNRLLLHGFIDENVAFAHTSILSLVRAKPYD 840
QY      841 LQIYPOERHSIRVPESGHEYLHLHYIQENIGSRIALKVI 882
Db      841 LQIYPOERHSIRVPESGHEYLHLHYIQENIGSRIALKVI 882

```

RESULT 3

```

US-10-170-789-38
; Sequence 38, Application US/10170789
; Publication No. US20030180930A1
; GENERAL INFORMATION:
; APPLICANT: Rachel E. Meyers
; APPLICANT: Olandt, Peter J.
; APPLICANT: Kapeller-Liberman, Rosana
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Williamson, Mark
; APPLICANT: Welch, Nadine
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,

```

```

; TITLE OF INVENTION: AND PROTEASE FAMILY MEMBERS AND USES THEREOF
; FILE REFERENCE: 10448-191001
; CURRENT FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 09/797,039
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06525
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/186,061
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 09/882,166
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19269
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212,078
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/934,406
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26052
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,740
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/861,801
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: PCT/US01/16549
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/205,508
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/801,267
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07138
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,454
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/829,671
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/US01/40483
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: US 60/197,508
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/961,721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235,023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045,367
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07074
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,420
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-170-789-38

Query Match      100.0%  Score 4700;  DB 12;  Length 882;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 882;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

; DB
; 61 AKAHPDFMVKRNDGDGSHDRITYLANSNGENRENTLFYSIPTINRAVIMLSWKEPL 120
; DLEOATLDVGMYSREBELREKRIGTVGIASVDYHOGSGFFLFOAGSGIYHVADGPGOG 180
; DLFQATLDVGMYSREBELREKRIGTVGIASVDYHOGSGFFLFOAGSGIYHVADGPGOG 180
; DLFQATLDVGMYSREBELREKRIGTVGIASVDYHOGSGFFLFOAGSGIYHVADGPGOG 180
; 181 FTQOPLRPMVLVETSCPNIRMDPKLCPADPDWIAFHSNDIWSNIVTREERLLTYVHNE 240
; 181 FTQOPLRPMVLVETSCPNIRMDPKLCPADPDWIAFHSNDIWSNIVTREERLLTYVHNE 240
; 241 ANNEEDASAGVATVLOEHPDYSYGMWCPAEFTPGSGKILRLYEENDESEIIVH 300
; 241 ANNEEDASAGVATVLOEHPDYSYGMWCPAEFTPGSGKILRLYEENDESEIIVH 300
; 301 TSPMLSTRADSFYKPTGTANPKYTFMSEIMIDAEGRIIDVIDKELIOFELIFEGVE 360
; 301 TSPMLSTRADSFYKPTGTANPKYTFMSEIMIDAEGRIIDVIDKELIOFELIFEGVE 360
; 361 YIARAGWTEGKYASILLDRSQTRLOIVLISPELFIPEVDVWERORLLISVPSYPL 420
; 361 YIARAGWTEGKYASILLDRSQTRLOIVLISPELFIPEVDVWERORLLISVPSYPL 420
; 421 IYEEETDIDIMINIDIFHFVPOSHHEEIEFTFASCKTGFRHLKYKITSILKSKYKRS 480
; 421 IYEEETDIDIMINIDIFHFVPOSHHEEIEFTFASCKTGFRHLKYKITSILKSKYKRS 480
; 481 GLPAPSDFCPIKEIATITSGEWEVLGRHSNIQDEVRLVYFEGTDSPLEHLLYVS 540
; 481 GLPAPSDFCPIKEIATITSGEWEVLGRHSNIQDEVRLVYFEGTDSPLEHLLYVS 540
; 541 YNPGSVTLTRDGRGSHSCISQHCDFISKYSNOKNHCVSILYLSPEDEPTCTKEF 600
; 541 YNPGSVTLTRDGRGSHSCISQHCDFISKYSNOKNHCVSILYLSPEDEPTCTKEF 600
; 601 WATILDSAGPLDYTPPELIFSFESCTGTLTGMLYKPHLDOPCKYPPVLYIYGPOVOL 660
; 601 WATILDSAGPLDYTPPELIFSFESCTGTLTGMLYKPHLDOPCKYPPVLYIYGPOVOL 660
; 661 VNNRFKGVYFRLNTLASIGYVVVIDNRGSCHRGLKEGAFKYGQIEIDDOVEGLO 720
; 661 VNNRFKGVYFRLNTLASIGYVVVIDNRGSCHRGLKEGAFKYGQIEIDDOVEGLO 720
; 721 LASRYDFIDLDRVGIHGSYGGYISLMALMORSIDIFRVAIAGAPVTLMIFYDTGTERYM 780
; 721 LASRYDFIDLDRVGIHGSYGGYISLMALMORSIDIFRVAIAGAPVTLMIFYDTGTERYM 780
; 781 GHPDQNGQYIYGSVAMQAEKFPSEPNRLLHLHGLDENVHFHATSILSFLVRGAKFYD 840
; 781 GHPDQNGQYIYGSVAMQAEKFPSEPNRLLHLHGLDENVHFHATSILSFLVRGAKFYD 840
; 841 LQIYQERHSIRVPESGEHYELHLHYLOENLGSRIALXVI 882
; 841 LQIYQERHSIRVPESGEHYELHLHYLOENLGSRIALXVI 882

RESULT 4
US-09-976-674-19
; Sequence 19, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Ol, Steve
; APPLICANT: Akinaanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Dunnen, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 19
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-19

```

```

Query Match      74.6%; Score 3504; DB 10; Length 658;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MAAMETEQGLVEIETADCEENIESODRPKLEPPYVERYSQKLLADTRKHYGM 60
DB 1 MAAMETEQGLVEIETADCEENIESODRPKLEPPYVERYSQKLLADTRKHYGM 60
QY 61 AKAPHEMFVKRNDPDPGHSRIYYLAMSNGENRENTLYSELPKTIINAAVILMSKPL 120
DB 61 AKAPHEMFVKRNDPDPGHSRIYYLAMSNGENRENTLYSELPKTIINAAVILMSKPL 120
QY 121 DLFOATLDYGMYSREBELRERKRI GTVGIASVDYHQSSTFLFOAGSGIYHVKGDPG 180
DB 121 DLFOATLDYGMYSREBELRERKRI GTVGIASVDYHQSSTFLFOAGSGIYHVKGDPG 180
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTREERRLTYVHNE 240
DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTREERRLTYVHNE 240
QY 241 ANMEDARSAGVATFVLQEEFDRYSGYWMCCKAETTPSGKILRLIYENDESEVEI 300
DB 241 ANMEDARSAGVATFVLQEEFDRYSGYWMCCKAETTPSGKILRLIYENDESEVEI 300
QY 301 TSPMLFTRRADSFRIYPTGTANPKYTFKMSIIMIDAGR11DIVIDKEI1QFEILLFEGVE 360
DB 301 TSPMLFTRRADSFRIYPTGTANPKYTFKMSIIMIDAGR11DIVIDKEI1QFEILLFEGVE 360
QY 361 YIARAGWTPBEGKYAMSIILDRSQTRLOIVLISPELFIPEDDVNERQLIESVDSVTP 420
DB 361 YIARAGWTPBEGKYAMSIILDRSQTRLOIVLISPELFIPEDDVNERQLIESVDSVTP 420
QY 421 IIVETTDIMINHDIFHVPQSHHEIEFI FASECTGFPHLYKITSILKESYKSSG 480
DB 421 IIVETTDIMINHDIFHVPQSHHEIEFI FASECTGFPHLYKITSILKESYKSSG 480
QY 481 GLPAPSPFKPIKEI1AITSGEWEVLGRHSGNIQVDEVRRLVYEGTQDSPLEHLLYVS 540
DB 481 GLPAPSPFKPIKEI1AITSGEWEVLGRHSGNIQVDEVRRLVYEGTQDSPLEHLLYVS 540
QY 541 YVNGEYTRLTDRGYSHSCCISOHCDFFISKYSNQKNPHCVSLYKLSPEDDPTCKTKEF 600
DB 541 YVNGEYTRLTDRGYSHSCCISOHCDFFISKYSNQKNPHCVSLYKLSPEDDPTCKTKEF 600
QY 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFITYG 655
DB 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFITYG 655

```

RESULT 5

```

US-09-976-674-11
; Sequence 11, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976, 674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11

```

```

; LENGTH: 661
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-11

```

```

Query Match      74.6%; Score 3504; DB 10; Length 661;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MAAMETEQGLVEIETADCEENIESODRPKLEPPYVERYSQKLLADTRKHYGM 60
DB 1 MAAMETEQGLVEIETADCEENIESODRPKLEPPYVERYSQKLLADTRKHYGM 60
QY 61 AKAPHEMFVKRNDPDPGHSRIYYLAMSNGENRENTLYSELPKTIINAAVILMSKPL 120
DB 61 AKAPHEMFVKRNDPDPGHSRIYYLAMSNGENRENTLYSELPKTIINAAVILMSKPL 120
QY 121 DLFOATLDYGMYSREBELRERKRI GTVGIASVDYHQSSTFLFOAGSGIYHVKGDPG 180
DB 121 DLFOATLDYGMYSREBELRERKRI GTVGIASVDYHQSSTFLFOAGSGIYHVKGDPG 180
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTREERRLTYVHNE 240
DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTREERRLTYVHNE 240
QY 241 ANMEDARSAGVATFVLQEEFDRYSGYWMCCKAETTPSGKILRLIYENDESEVEI 300
DB 241 ANMEDARSAGVATFVLQEEFDRYSGYWMCCKAETTPSGKILRLIYENDESEVEI 300
QY 301 TSPMLFTRRADSFRIYPTGTANPKYTFKMSIIMIDAGR11DIVIDKEI1QFEILLFEGVE 360
DB 301 TSPMLFTRRADSFRIYPTGTANPKYTFKMSIIMIDAGR11DIVIDKEI1QFEILLFEGVE 360
QY 361 YIARAGWTPBEGKYAMSIILDRSQTRLOIVLISPELFIPEDDVNERQLIESVDSVTP 420
DB 361 YIARAGWTPBEGKYAMSIILDRSQTRLOIVLISPELFIPEDDVNERQLIESVDSVTP 420
QY 421 IIVETTDIMINHDIFHVPQSHHEIEFI FASECTGFPHLYKITSILKESYKSSG 480
DB 421 IIVETTDIMINHDIFHVPQSHHEIEFI FASECTGFPHLYKITSILKESYKSSG 480
QY 481 GLPAPSPFKPIKEI1AITSGEWEVLGRHSGNIQVDEVRRLVYEGTQDSPLEHLLYVS 540
DB 481 GLPAPSPFKPIKEI1AITSGEWEVLGRHSGNIQVDEVRRLVYEGTQDSPLEHLLYVS 540
QY 541 YVNGEYTRLTDRGYSHSCCISOHCDFFISKYSNQKNPHCVSLYKLSPEDDPTCKTKEF 600
DB 541 YVNGEYTRLTDRGYSHSCCISOHCDFFISKYSNQKNPHCVSLYKLSPEDDPTCKTKEF 600
QY 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFITYG 655
DB 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFITYG 655

```

RESULT 6

```

US-09-976-674-7
; Sequence 7, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976, 674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 690

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-7

```

```

Query Match      74.6%; Score 3504; DB 10; Length 690;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MAAMETEQLGVEIFETADCEENIESODRPLKEFFYVERYSQKLLADTRKHYGM 60
DB 1 MAAMETEQLGVEIFETADCEENIESODRPLKEFFYVERYSQKLLADTRKHYGM 60
QY 61 AKAPHEMFVVRNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTINRAAVLMSWKPL 120
DB 61 AKAPHEMFVVRNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTINRAAVLMSWKPL 120
QY 121 DLFOATLDYGMYSSEBELRERKRI GTVGASDYHOGSGTFLEQAGSGIYHVDGPG 180
DB 121 DLFOATLDYGMYSSEBELRERKRI GTVGASDYHOGSGTFLEQAGSGIYHVDGPG 180
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHNSNDIWSNIYTRERRLTYVHNE 240
DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHNSNDIWSNIYTRERRLTYVHNE 240
QY 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTSQGGKILRIEENDESEVEI 300
DB 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTSQGGKILRIEENDESEVEI 300
QY 301 TSPMLETRRADSRFPYPTGTANPKVTKMEIMDAGR11DIVDKELIQPELLEFEGVE 360
DB 301 TSPMLETRRADSRFPYPTGTANPKVTKMEIMDAGR11DIVDKELIQPELLEFEGVE 360
QY 361 YIARAGWTPRGKXAMSLILDRSQTRQIVLISELFIPEVDDWVERORLIESVDSV 420
DB 361 YIARAGWTPRGKXAMSLILDRSQTRQIVLISELFIPEVDDWVERORLIESVDSV 420
QY 421 IYEEETDWINIHDIHFVFPQSHHEIEFI FASECKTGRRLYKITSILKESYKSSG 480
DB 421 IYEEETDWINIHDIHFVFPQSHHEIEFI FASECKTGRRLYKITSILKESYKSSG 480
QY 481 GLPAPSDFKPIKEEIAITSGEWEVIGRHGNSIQVEVRRLVYFEGTKDSPLBHL 540
DB 481 GLPAPSDFKPIKEEIAITSGEWEVIGRHGNSIQVEVRRLVYFEGTKDSPLBHL 540
QY 541 YVNPGEVTRLTDGYSHSCCISQCHDFISKYSNQNPKHCVSLYKSSPEDDPCTK 600
DB 541 YVNPGEVTRLTDGYSHSCCISQCHDFISKYSNQNPKHCVSLYKSSPEDDPCTK 600
QY 601 MATILDSAGLDPYTPPEIFSFESTTGFTLYGMLYKPHDLPQKXPTVLFIVG 655
DB 601 MATILDSAGLDPYTPPEIFSFESTTGFTLYGMLYKPHDLPQKXPTVLFIVG 655

```

```

RESULT 7
US-09-976-674-21
; Sequence 21, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akimsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Homo sapiens

```

```

; ORGANISM: Homo sapiens
US-09-976-674-21

```

```

Query Match      68.9%; Score 3236; DB 10; Length 613;
Best Local Similarity 100.0%; Pred. No. 1,4e-301;
Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MAAMETEQLGVEIFETADCEENIESODRPLKEFFYVERYSQKLLADTRKHYGM 60
DB 1 MAAMETEQLGVEIFETADCEENIESODRPLKEFFYVERYSQKLLADTRKHYGM 60
QY 61 AKAPHEMFVVRNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTINRAAVLMSWKPL 120
DB 61 AKAPHEMFVVRNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTINRAAVLMSWKPL 120
QY 121 DLFOATLDYGMYSSEBELRERKRI GTVGASDYHOGSGTFLEQAGSGIYHVDGPG 180
DB 121 DLFOATLDYGMYSSEBELRERKRI GTVGASDYHOGSGTFLEQAGSGIYHVDGPG 180
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHNSNDIWSNIYTRERRLTYVHNE 240
DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHNSNDIWSNIYTRERRLTYVHNE 240
QY 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTSQGGKILRIEENDESEVEI 300
DB 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTSQGGKILRIEENDESEVEI 300
QY 301 TSPMLETRRADSRFPYPTGTANPKVTKMEIMDAGR11DIVDKELIQPELLEFEGVE 360
DB 301 TSPMLETRRADSRFPYPTGTANPKVTKMEIMDAGR11DIVDKELIQPELLEFEGVE 360
QY 361 YIARAGWTPRGKXAMSLILDRSQTRQIVLISELFIPEVDDWVERORLIESVDSV 420
DB 361 YIARAGWTPRGKXAMSLILDRSQTRQIVLISELFIPEVDDWVERORLIESVDSV 420
QY 421 IYEEETDWINIHDIHFVFPQSHHEIEFI FASECKTGRRLYKITSILKESYKSSG 480
DB 421 IYEEETDWINIHDIHFVFPQSHHEIEFI FASECKTGRRLYKITSILKESYKSSG 480
QY 481 GLPAPSDFKPIKEEIAITSGEWEVIGRHGNSIQVEVRRLVYFEGTKDSPLBHL 540
DB 481 GLPAPSDFKPIKEEIAITSGEWEVIGRHGNSIQVEVRRLVYFEGTKDSPLBHL 540
QY 541 YVNPGEVTRLTDGYSHSCCISQCHDFISKYSNQNPKHCVSLYKSSPEDDPCTK 600
DB 541 YVNPGEVTRLTDGYSHSCCISQCHDFISKYSNQNPKHCVSLYKSSPEDDPCTK 600
QY 601 MATILDS 607
DB 601 MATILDS 607

```

```

RESULT 8
US-09-976-674-3
; Sequence 3, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akimsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 863
; TYPE: PRT
; ORGANISM: Homo sapiens

```


RESULT 10
US-09-976-674-27
; Sequence 27, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 892
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-976-674-27

Query Match 61.1%; Score 2870; DB 10; Length 892;
Best Local Similarity 61.5%; Pred. No. 4e-266;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;
QY 35 FVYERYSMQKLKLLADTRKHYGMMAKAPHDPMFYKRNDDPGPHSDRIYILAMSGENR 94
DB 53 FOVQKSHMDGLRSIIHSGSRKYSGLIVNKAHPDFQVQKTDGSGPHSHRIYILGMPYGSRE 112
QY 95 NTLFYSEIPKTIINRAVAVLMSKPLLDLQATLDYGMYSREBELLRERKRIQVGTIASYD 154
DB 113 NSLTVSEIPKTKRKALLLSWKQMLDHPQATPHHGVSREBELLRERKRIQVGTIASYD 172
QY 155 YHOGSGTFLFOAGSGIYHVKGDPGFTQOPLRPMLVETSCNINMDPKLCPADPDMIAF 214
DB 173 FHSSEGLFLFOASNSLIFCRDGGKNGFMVSPKPLEIKTQCGSPRMDPKICPADPAFSPF 232
QY 215 IHSNDIWSNITVREERLTYVHNLAMNEEDARASAGVATFVLOEFPDRYSGYMWCPKAE 274
DB 233 INNSDLWYANITETGERRLTFCHQGLSNVLDPKSAGVATFVLOEFPDRYSGYMWCPKAE 292
QY 275 TTPSGG-KILRLIYENDESEVEIIVHTSPMLSTRADSFYPKGTGANPKVTPKMSIEM 333
DB 293 WEGSGGLKTLRLIYEVDESEVEIIVHSPALERKTSYRPRKGSKNPKIALKLAEPQ 352
QY 334 IDAEGRIIDVDELIOPELIFEGVEYIARAGWTPCKGKAMSIILDSQTRLQIVLISP 393
DB 353 TDSQKIVSTQEKELVQFSSLPFVVEYIARAGWTRDQKIYAMWFLDRPQWLQVLLP 412
QY 394 ELFIPEDDVNERQLIESVDSVTPPLIYEETDININHDIFHVFQSH-EEIEIFL 452
DB 413 ALFIPTSTNEBOKLASARAVERNQPVYVEVTVWVNHVDFPFPQSEEDLCLRL 472
QY 453 ASECKTGRRLIYKTSILKSKYRSOGGLPAPSPKCPICEKIALITGSEWVLCRHSN 512
DB 473 ANECKTGFCHLYKTVAVLKSGYDMSPEPSPGDEDFKCPICEKIALITGSEWVLCRHSN 532
QY 513 IQVDEVRRLVYFEGTKDPSLHHLVYVSYVNGEYVTRLTDRGYSHSCCISQCHDFPISKY 572
DB 533 IWNVEETLVYFQGTOKDPLHHLVYVSYVNGEYVTRLTDRGYSHSCCISQCHDFPISKY 592
QY 573 SNQKNPHVSLYKLSPEDDPTCKTEWATILDSAGLPRYTPPEIFSESTGTFTLYG 632
DB 593 SSVSTPVCVHYKLSGPDDELHQPREFWASMEALASCPDYVPEIIFHFTSRDVLRYG 652
QY 633 MLVPHDLOPKKPYLVLFYIGPOVOYLWNRFKGVKFERINTASLGIVVYVINDRSGC 692
DB 653 MIVPHALOPKPKPYLVLFYIGPOVOYVWNSFKGIKTLRLNTASLGIVVYVINDRSGC 712
QY 693 HRLKFGAFLYKXNGQIEIDDQVEGLQYLASRYDFIDLRVGIHGWSYGYLSIMALMQR 752

DB 713 QRGIRREGALKNMGQVEIDQVEGLQVFAEKYGFIDLSRAVILHGWSYGYLSIMALMQR 772
QY 753 SDIFRYALAGAPVLMITFYDTGTERYMGHPRDQEGYILGSVAMQKPEPSENNLL 812
DB 773 POFKVALIAGAPVLMWADVTGTERYMDPENNQGYEAGVALHLEKLPENBNRLIL 832
QY 813 HGFLENVHFAHTSILSLFVRAKPEYDLOIYQOERSIRIVPSSGEYELHLHYLOENL 872
DB 833 HGFLENVHFAHTSILSLFVRAKPEYDLOIYQOERSIRIVPSSGEYELHLHYLOENL 892

RESULT 11
US-09-976-674-33
; Sequence 33, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 879
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-976-674-33
Query Match 60.0%; Score 2820.5; DB 10; Length 879;
Best Local Similarity 60.7%; Pred. No. 2.2e-261;
Matches 510; Conservative 132; Mismatches 183; Indels 15; Gaps 3;
QY 35 FVYERYSMQKLKLLADTRKHYGMMAKAPHDPMFYKRNDDPGPHSDRIYILAMSGENR 94
DB 53 FOVQKSHMDGLRSIIHSGSRKYSGLIVNKAHPDFQVQKTDGSGPHSHRIYILGMPYGSRE 112
QY 95 NTLFYSEIPKTIINRAVAVLMSKPLLDLQATLDYGMYSREBELLRERKRIQVGTIASYD 154
DB 113 NSLTVSEIPKTKRKALLLSWKQMLDHPQATPHHGVSREBELLRERKRIQVGTIASYD 172
QY 155 YHOGSGTFLFOAGSGIYHVKGDPGFTQOPLRPMLVETSCNINMDPKLCPADPDMIAF 214
DB 173 FHSSEGLFLFOASNSLIFCRDGGKNGFMVSPKPLEIKTQCGSPRMDPKICPADPAFSPF 232
QY 215 IHSNDIWSNITVREERLTYVHNLAMNEEDARASAGVATFVLOEFPDRYSGYMWCPKAE 274
DB 233 INNSDLWYANITETGERRLTFCHQGLSNVLDPKSAGVATFVLOEFPDRYSGYMWCPKAE 292
QY 275 TTPSGG-KILRLIYENDESEVEIIVHTSPMLSTRADSFYPKGTGANPKVTPKMSIEM 333
DB 293 WEGSGGLKTLRLIYEVDESEVEIIVHSPALERKTSYRPRKGSKNPKIALKLAEPQ 352
QY 334 IDAEGRIIDVDELIOPELIFEGVEYIARAGWTPCKGKAMSIILDSQTRLQIVLISP 393
DB 353 TDSQKIVSTQEKELVQFSSLPFVVEYIARAGWTRDQKIYAMWFLDRPQWLQVLLP 412
QY 394 ELFIPEDDVNERQLIESVDSVTPPLIYEETDININHDIFHVFQSH-EEIEIFL 452
DB 413 ALFIPTSTNEBOKLASARAVERNQPVYVEVTVWVNHVDFPFPQSEEDLCLRL 472
QY 453 ASECKTGRRLIYKTSILKSKYRSOGGLPAPSPKCPICEKIALITGSEWVLCRHSN 512
DB 473 ANECKTGFCHLYKTVAVLKSGYDMSPEPSPGDEDFKCPICEKIALITGSEWVLCRHSN 531
QY 513 IQVDEVRRLVYFEGTKDPSLHHLVYVSYVNGEYVTRLTDRGYSHSCCISQCHDFPISKY 572


```

Db      532 ----- KGKDTPLFHHLLVYVVEAAEIEYRLTRTPGSHSCMSQNFDMFVSHY 579
QY      573 SMOKNPHCSLYLKSSPEDDPTCKTEPMATILDSAGRLPDYTPPEIFSESTGTFLYG 632
Db      580 SVSTSPCCYHVYKLSGDDDDPLHKORFWASMEASCPEDYVPPEIFPHETRSDDLRYG 639
QY      633 MLYKPHDLOPGKKYPLVLFYGGPQVOLVNNRFGKYFFELNTLASIGYVVVVYIDRGSC 692
Db      640 MLYKPHALOPGKKHPTVLFPYGGPQVOLVNNRFSKIKYLRLNTLASLGAVVYIDRGSC 699
QY      693 HRGLPEEGFKYKMGQIEIDQVEGLOYLASRDPLDLDVGVGHGMSYGGYLSLMAIMOR 752
Db      700 QRLGTFEGALKMQMGVEILEDQVEGLOFVAEKGFILDSRVALHGMYSYGGFLSLMGILHK 759
QY      753 SDIFPVALAGAEVTLWIFPYDTGYTERVMYMHPRQNEOGYILGSAVMAOEKPESEPNLLIL 812
Db      760 POFVKVIALAGAVTVMAAYDTGYTERYMDVPENNQGYEAGSVALAVEKLPENPNLLIL 819
QY      813 HGFLEUENHFAHTSLISFLVRAAGKYDQIYIPQENHSTRVPSGSHYELAHLLHYQENT 872
Db      820 HGFLEUENHAFPTINELVSOILIRGAKYQLOIYIPNEHSHSTRCPSSGSHYEVTLLHFLQEVY 879

```

```

RESULT 12
US-09-976-674-35
: Sequence 35, Application US/09976674
: Patent No. US20020115843A1
: GENERAL INFORMATION:
: APPLICANT: Qi, Steve
: APPLICANT: Akinsanya, Karen
: APPLICANT: Riviere, Pierre
: APPLICANT: Junien, Jean-Louis
: TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
: FILE REFERENCE: 70669
: CURRENT APPLICATION NUMBER: US/09/976,674
: CURRENT FILING DATE: 2001-10-12
: PRIOR APPLICATION NUMBER: US 60/240,117
: PRIOR FILING DATE: 2000-10-12
: NUMBER OF SEQ ID NOS: 61
: SOFTWARE: SeqId version 3.1
: SEQ ID NO 35
: LENGTH: 879
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-976-674-35

```

Query Match	60.0%;	Score 2820.5;	DB 10;	Length 879;
Best Local Similarity	60.7%;	Pred No. 2.2e+261;		
Matches 510;	Conservative 183;	Indels 15;	Gaps 3	

QY	35	FYVRYSWSSQCKLLADLRKXHYGMMAAPHDPMVKKNDPDGPHSDIYYIYLAAMGENRE	94
Db	53	FQYKHSMDGLRSIIHSRKSXSGLIYNAKPHDFQVQKTDSEGPSSHLYIYGMVGSSE	112
QY	95	NTLFYSELPKTIINPAVLMSKPYLLDLPATLDYGMYSREBELLRERRIGTVIASYD	154
Db	113	NSLIYSELPKKVYRREALIILMSKMDLDFQATPHHGVYSREBELLRERRKLGVFITSYD	172
QY	155	YHQSGETFLPQASGIIYHVKDQSGPGGFPQOPLRPLVYTSQDNTRMDEKLCPADDDMAF	214
Db	173	FHSRSGFLPQASLSLPHCRDQKNGKPFVWSFMKPLEITQCSGPRMDEKLCPADAPFSF	232
QY	215	IHSMDIWSNVTVEERRLIYVENELANNEEDASAGYATVLOEFNDRYSGYWWCPYAE	274
Db	233	INNSDLMVANIETGEERRLTFCHOGLSVNLDDPSAGATAVTIOEFPDRFTGYWVWCPYAS	292
QY	275	TTPSGG-KIIRILYEENDESEVEIIHYVTSPMLEFRADSFYPKGTANPVTFMFSIM	333
Db	293	WESGEGKTLRIILEYDESEVEVIHNPSPALBERKIDSYYPRPTGSNPKALKLKAEPQ	352
QY	334	IDABGRIDVIDXELIOPFELLFPGCYVIAAGMTBEGKAMSIILDRSQRLQIVLISP	393
Db	353	TDSQGRKVSIOEKELVOPFSGSLFKEVETIAAGTTRQKAMAMFELRPOOMLQIVLILP	412

[illegible]

```

RESULT 13
US-09-976-674-29
; Sequence 29, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Unien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPIP
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 832
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-976-674-29

```

Query Match	51.2%	Score 2406;	DB 11;	Length 832;
Best local similarity	59.0%	Pred. NO. 1.4e-22;		
Matches 434;	Conservative 122;	Mismatches 177;	Indels 2;	Gaps 2

QY 35 FVYERYSMSOGLKLLDLADREYHGMMAKAHHDFVRYKRPDPDGHSDRIYYILAMSGENRE 94
 Db 53 FOYQKSWMDLKRSTIHGSRYSGLIYNKAPADFOYQVKTRESGPHSHLYYLGMPSYGRE 112
 QY 95 NTLFYSEIPIKTRIRAAVLMYLMKRPDLDPFOATLDYGYMSREELIREKRIYGTGIAASYD 154
 Db 113 NSLLYSEIPIKRVKREKALLLSMKOMLDHFOATPHHGIVYSREELIREKRIYGVGITSYD 172
 QY 155 YHOGSGTFLEFOAGSGIYHVVDGPOGFTOOLPRNLVETSCPNIRMDPKLCPADPDWIAF 214
 Db 173 FHSESGTFLEFOAGSNLSLFCRCDGCKGNGPMYSPMKPDLKIKTQCSGRMDPKICPADPAFSEF 232

QY 215 IHSNDIWSNVTREBRRLTYVHNELANMEDASAGVATFVLQEEBDRYSGYWCCKAE 274
 DB 233 INNSDLWANIETGERRLTFCHQGLSNVLDPKSAGVATFVLQEEBDRFTGVMCPYAS 292
 QY 275 TTSGG-KILRIYEENDESEVEIIVHVSPLMELTRRADSPRYPKGTANPKVFKMSEIM 333
 DB 293 WESSEGLKTLRIIYEVEDESEVEIIVHVSPLMELTRRADSPRYPKGTANPKVFKMSEIM 352
 QY 334 IDAEGRIIDVIDKEILOPFELFEGVEYIARAGMTBEGKANSILDRSOTRLQYVLISP 393
 DB 353 TDSQGIKIVSTQEKELVQFSSLPFKVEYIARAGMTBEGKANSILDRSOTRLQYVLISP 412
 QY 394 ELFIPEVDVMEORQLIESVPDSVTPLLIYEETIDMINIHDIHFVPOSH-EEIEFTF 452
 DB 413 ALFIPTSTENEBORLASAAVPRNQPYVVEVTNWINVDIYFPFGSGEDELCTLR 472
 QY 453 ASECKTGFRLHYKITSILKESKYRSSGGLPAPSDFKCPIKEBIAITSGMEVLAGHGSN 512
 DB 473 ANECKTGFCCHLYKVTAVLKSQYDWSPEFSGEDEFKCPIKEBIAITSGMEVLAGHGSN 532
 QY 513 IOYDEVRLVYFEGTKDPLSHHLVYVSVNPGEVTRLTDGYSHSCCISQHCDFISKY 572
 DB 533 IWNVEETLVYFQGTDPLEHHLVYVSEYAEAGIYVRLTTPGFSHSCMSQNFDFVSHY 592
 QY 573 SNOKNPHCVSLYKLSPEDDPTCKTKEFMATILDSAGPLDYPPEIFSFESTTGFTLYG 632
 DB 593 SSVSTPCVHYKLSGPDDDLHKQRFPMASMEASCPDYVPEIHFHTRSDVRLYG 652
 QY 633 MLKPHDLOPGKRYPTVLYFGGPOVULVNNRFGVKYFRANTLASLGVVYVINDRSGC 692
 DB 653 MIYKPHALOFGKHPVLYFGGPOVULVNNRFGVKYFRANTLASLGVVYVINDRSGC 712
 QY 693 HRGLKEGAFKRYKMGQIIEIDQVEGLQYLASRYDFIDLRVGIHGSYGYLSIMALMOR 752
 DB 713 QRGIRFEGALKNQMGVIEIDQVEGLQYFAKXGFIIDLSVALHGSYGYLSIMGLINK 772
 QY 753 SDIFRVAIAGAPVTL 767
 DB 773 PQVFKAOPLAYPRL 787
 RESULT 14
 US-09-976-674-31
 ; Sequence 31, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Qi, Steve
 ; APPLICANT: Akusanya, Karen
 ; APPLICANT: Riviere, Pierre
 ; APPLICANT: Junien, Jean-Louis
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
 ; FILE REFERENCE: 70669
 ; CURRENT APPLICATION NUMBER: US/09/976,674
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: US 60/240,117
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 31
 ; LENGTH: 832
 ; TYPE: PRP
 ; ORGANISM: Homo sapiens
 US-09-976-674-31

Query Match 51.2%; Score 2406; DB 10; Length 832;
 Best Local Similarity 59.0%; Pred. No. 1,4e-221;
 Matches 434; Conservative 122; Mismatches 177; Indels 2; Gaps 2;
 QY 35 FYEERYGMSQKLLADTRKYHGYMAKAPHDPMFVGRNDPDPGSHDRIYLLMSENGE 94
 DB 53 FOYQKSHMDGRLSRILHSGRKYSGIIVNKAHPDFQVQKDSGPHRLYLLMPPGSR 112
 QY 95 NTLFYSEIPTINPAVALMSKPLDLDFOATIDYGMYSREBELRKRKRGITGVIAASYD 154

DB 113 NSLIXSEIPKVRKREALLLSMKOMLDFQATPHHGVYSREBELRKRKRGITGVIAASYD 172
 QY 155 YHQSGLFLOAGSGIYHVKDGPOGFTQOPLRPNLVETSCNIRMDPKLCPADPDMAIF 214
 DB 173 FHSSEGLFLFOASNLIFCHRDGKNGFVNSPMKLEIKTQCSGRMPDKICPADPAFFSF 232
 QY 215 IHSNDIWSNVTREBRRLTYVHNELANMEDASAGVATFVLQEEBDRYSGYWCCKAE 274
 DB 233 INNSDLWANIETGERRLTFCHQGLSNVLDPKSAGVATFVLQEEBDRFTGVMCPYAS 292
 QY 275 TTSGG-KILRIYEENDESEVEIIVHVSPLMELTRRADSPRYPKGTANPKVFKMSEIM 333
 DB 293 WESSEGLKTLRIIYEVEDESEVEIIVHVSPLMELTRRADSPRYPKGTANPKVFKMSEIM 352
 QY 334 IDAEGRIIDVIDKEILOPFELFEGVEYIARAGMTBEGKANSILDRSOTRLQYVLISP 393
 DB 353 TDSQGIKIVSTQEKELVQFSSLPFKVEYIARAGMTBEGKANSILDRSOTRLQYVLISP 412
 QY 394 ELFIPEVDVMEORQLIESVPDSVTPLLIYEETIDMINIHDIHFVPOSH-EEIEFTF 452
 DB 413 ALFIPTSTENEBORLASAAVPRNQPYVVEVTNWINVDIYFPFGSGEDELCTLR 472
 QY 453 ASECKTGFRLHYKITSILKESKYRSSGGLPAPSDFKCPIKEBIAITSGMEVLAGHGSN 512
 DB 473 ANECKTGFCCHLYKVTAVLKSQYDWSPEFSGEDEFKCPIKEBIAITSGMEVLAGHGSN 532
 QY 513 IOYDEVRLVYFEGTKDPLSHHLVYVSVNPGEVTRLTDGYSHSCCISQHCDFISKY 572
 DB 533 IWNVEETLVYFQGTDPLEHHLVYVSEYAEAGIYVRLTTPGFSHSCMSQNFDFVSHY 592
 QY 573 SNOKNPHCVSLYKLSPEDDPTCKTKEFMATILDSAGPLDYPPEIFSFESTTGFTLYG 632
 DB 593 SSVSTPCVHYKLSGPDDDLHKQRFPMASMEASCPDYVPEIHFHTRSDVRLYG 652
 QY 633 MLKPHDLOPGKRYPTVLYFGGPOVULVNNRFGVKYFRANTLASLGVVYVINDRSGC 692
 DB 653 MIYKPHALOFGKHPVLYFGGPOVULVNNRFGVKYFRANTLASLGVVYVINDRSGC 712
 QY 693 HRGLKEGAFKRYKMGQIIEIDQVEGLQYLASRYDFIDLRVGIHGSYGYLSIMALMOR 752
 DB 713 QRGIRFEGALKNQMGVIEIDQVEGLQYFAKXGFIIDLSVALHGSYGYLSIMGLINK 772
 QY 753 SDIFRVAIAGAPVTL 767
 DB 773 PQVFKAOPLAYPRL 787
 RESULT 15
 US-09-976-674-37
 ; Sequence 37, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Qi, Steve
 ; APPLICANT: Akusanya, Karen
 ; APPLICANT: Riviere, Pierre
 ; APPLICANT: Junien, Jean-Louis
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
 ; FILE REFERENCE: 70669
 ; CURRENT APPLICATION NUMBER: US/09/976,674
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: US 60/240,117
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 37
 ; LENGTH: 819
 ; TYPE: PRP
 ; ORGANISM: Homo sapiens
 US-09-976-674-37

Query Match 50.1%; Score 2356.5; DB 10; Length 819;
 Best Local Similarity 58.1%; Pred. No. 7.7e-217;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:08:55 ; Search time 23.176 Seconds
(without alignments)
3659.853 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700

Sequence: 1 MAAMMETEQLCVEIFETADDC.....HLHLYQENIGSRIALKVI 882

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922.5	19.6	931	T32919	hypothetical prote
2	842	17.9	738	A87516	dipeptidyl peptid
3	754	16.0	741	JC5142	X-Pro dipeptidyl-P
4	620	13.2	711	S66261	X-Pro dipeptidyl-P
5	529	11.3	766	CDH026	dipeptidyl-peptid
6	519.5	11.1	793	T41703	dipeptidyl aminope
7	518.5	11.0	760	S23752	dipeptidyl-peptid
8	510.5	10.9	792	A39914	dipeptidyl-peptid
9	484.5	10.3	818	A30107	dipeptidyl aminope
10	465	9.9	931	A49737	dipeptidyl aminope
11	451.5	9.6	779	T25173	hypothetical prote
12	445.5	9.5	779	T25173	hypothetical prote
13	440	9.4	803	A41793	dipeptidyl aminope
14	427	9.1	803	I54331	dipeptidyl aminope
15	427	9.1	865	I54331	dipeptidyl aminope
16	378	8.0	829	T19514	dipeptidyl aminope
17	374	8.0	795	F82958	hypothetical prote
18	331.5	7.1	759	I38593	dipeptidyl-peptid
19	293	6.2	743	T37700	fibroblast activat
20	282	6.0	642	C71137	probable dipeptid
21	269.5	5.7	631	H75007	hypothetical prote
22	268	5.7	683	E87495	prolyl oligopeptid
23	255.5	5.4	709	B82580	alanyl dipeptidyl
24	249.5	5.3	657	E70025	hypothetical prote
25	236	5.0	632	F71174	hypothetical prote
26	223.5	4.8	632	E75057	hypothetical prote
27	217.5	4.6	536	F90299	acylaminoacyl-pept
28	200.5	4.3	591	H72474	probable acylamino
29	185	3.9	569	S74053	probable acylamino

30	183.5	3.9	608	2	F83397	probable peptidase
31	182.5	3.9	674	2	B84381	acylaminoacyl-pept
32	180.5	3.8	721	2	T09631	probable acylamino
33	180	3.8	667	2	A87711	prolyl oligopeptid
34	172	3.7	732	1	U04655	dipeptidyl aminope
35	171	3.6	732	1	S07624	acylaminoacyl-pept
36	170.5	3.6	572	2	F72455	acylaminoacyl-pept
37	163.5	3.5	659	2	F72568	probable acylamino
38	161	3.4	676	2	C97775	acylaminoacyl-pept
39	160.5	3.4	732	1	JU0132	acylaminoacyl-pept
40	160.5	3.4	745	2	T33751	hypothetical prote
41	158	3.4	598	2	F84199	hypothetical prote
42	157	3.3	614	2	E75094	peptidase (impor
43	155.5	3.3	654	2	AD3183	hypothetical prote
44	155	3.3	629	2	T15945	hypothetical prote
45	154.5	3.3	828	2	G87564	hypothetical prote

ALIGNMENTS

RESULT 1

T32919

hypothetical protein K02F2.1 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 29-Oct-1999, #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C/Accession: T32919

R/Maggi, L.; Goela, D.

submitted to the EMBL Data Library, January 1998

A/Description: The sequence of C. elegans cosmid K02F2.

A/Reference number: Z21246

A/Accession: T32919

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1931 <MAG>

A/Cross-references: EMBL:AF043699; PIN:AA97564.1; GSPDB:GN00019; CESP:K02F2.1

A/Experimental source: strain Bristol N2; clone K02F2

C/Genetics:

A/Map position: 1

A/Introns: 58/3; 82/2; 131/2; 178/2; 275/3; 322/2; 404/2; 441/1; 464/2; 486/3; 528/2; 5

Query Match	19.6%	Score 922.5;	DB 2;	Length 931;
Best Local Similarity	29.3%;	Pred. No. 1.4e-57;		
Matches	282;	Conservative 145;	Mismatches 326;	Indels 211; Gaps 41;
QY	33	EPFVEXYSQKLLADIRKYGMMAKAPHDPMFVKRNDPQPSDRITYLAMSGEN	92	
DB	36	EPARFETRSFQ--LIDHARSWKTEVRGMTQGFYKISIMRAE--KDRIMVYAISSVP	89	
QY	93	RENT--LFYSEIP-KTINRAAVLMLSWKPLDLFOATLDVGMV-----SRE	135	
DB	90	GTNQSIFSVTIPLELVEKAQVADRK-----FELTKSGYVNDSTIRMSCKRTPPSAE	142	
QY	136	EELIREKRIGTV-GIASYDYGQ--SGTFLPQAG-----	167	
DB	143	FTLQCEHRSQVVTGIDYEIRNGKMTLMGDDQFRNPLNEALATPIAVPDQSSTEP	202	
QY	168	-----SGTHVKQGGQ--FTQQLRPMLVETSCP-----NIRMDPLCP	206	
DB	203	MDISEGSIITGKGCSEAPQSSVTPVTPRIPKPTTSIEKFPATPNNFVSAAVCP	262	
QY	207	ADPDWIAFIHNSINDINITYRREARRLTYYHNELANNEEDARSGVATFVLOEFDRYSG	266	
DB	263	ADSSLLAVLVANKQYI-----EKNGKTIHRSSNSKHITN--GVPSYIVQELERPEG	313	
QY	267	YWCCKAETTPSGGKILRIIYENDESEVEIHHV-----TSPMLETRRADSRFRYKXTG	319	
DB	314	IWM-SESKT-----RLTYHVNBEKVAESQGVGVDPVAPM-----KYRAG	355	
QY	320	TANPKVTFKMSLIMDAEGRILIDVDEKLIQPEILPEG--VEYIRAGWTPEGKXAMS	376	
DB	356	TKNAYSTLRN---VILENGKAYDVPKLD-----EVIYKHPFYEYITRAGFSDGTVMV	407	

```

QY 377 ILDRSOTRLQIVLI-----SPELFIVYEDDVMERORLIESVP 414
DB 408 QWMSDQAOCCILLIPYDPLPRELGSIKEDNIQSTDLNMGWMD-KSHEETMEKRP 466
QY 415 DSVT--PLIYBETTDVININHDIFHVPQSHHEE--IEFIASECKTGFRHLKYTSIL 470
DB 467 RGLKRGTVQIHKARNDYMINFHAIYPLKITDEHPMEFYCLEKPNG-SCLALISLML 525
QY 471 KESKYKSSGGLPASPDPKCPRIKEIATSGMEVTLGSHGNSIQVDERRLVYEGTODS 530
DB 526 DONGT-----CRHTEKLLMAENFSI--NKSNGIIVDEVREBELVYVYANESH 569
QY 531 PLEHLLVYVSVNPEGEVRLTLDRGYSHSCISQHCDFISKYSNOK-----NPHCVSLYKL 586
DB 570 PLEWMI-CVSHYRQGHQJTESGI-----C-FKSRBRANKLALDDHGFACVMT 617
QY 587 S--SPEDDPTCKTER-N-----ATLIDSAGP-LPD--YTPPELFSSES--TTG 627
DB 618 SVGSPAE---CRFYSFRWKENEVLPSTVYANITVSGHPQDPDLHFDSPEMIBFQSKTG 674
QY 628 FTLGMLYKPHDLOPGKKYPTVLFYGGQVOIVNNRKYKVPRLNTLASLGYVVVID 687
DB 675 LMAYAMIRPSNPDYKIPVPHYVYGGPQIVHNDSTWQYLR---FCRLGYVVVID 731
QY 688 NRGSCHRGKFEKGAFTKMMQGEIDQVEGLQYLASRY-DFIDLDVYGHGMSYGYLSTL 746
DB 732 NRGSARHGIIEFERHIIKMGITVEVEDQVEGLQMLAERTGPMDSRVVHVMSSYGGTAL 791
QY 747 MALMORSDIFRVALAGAPVTIMITYDTGYTERYMGHPDQNGYLLGSVMAQAEKFPSEP 806
DB 792 QMAKHPIYRAALAGAVSDMRLYDPAITERVMGYP--LEEHVYGASSITGLVEKLPDEP 850
QY 807 NRLLILHGFDENHFAHSTILSLFLVAGKPYDLOIYPOERHSIRPESHEHLEHLLH 866
DB 851 NRMLVHVGMDENHFAHLEHVDCEIKKGMHLEVIFFENRHHVNNDASTYLDARMY 910
QY 867 YLQGE 870
DB 911 PAQO 914

RESULT 2
A87516
dipeptidyl peptidase IV (imported) - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: A87516
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.;
B.J. Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.;
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: A87516
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-738 <STO>
A/Cross-references: GB:AE005673; NID:g14323647; PIDN:AAK2125.1; GSPDB:GN00148
C/Genetics:
A/Gene: CC2154

Query Match 17.9%; Score 842; DB 2; Length 738;
Best Local Similarity 29.9%; Pred. No. 5.5e-52;
Matches 221; Conservative 115; Mismatches 262; Indels 141; Gaps 24;

QY 133 SRREELRKRRTIGVGIASVDYHQSCTFLFQAGSGIT--HYVDGPGQGTQOPLFRL 190
DB 94 SBAEKARBRARVASARIVERYSWDQGRFLIVPLDGLDLYDAVADGKITRLTLE----- 146

QY 191 VETSCPNIRMDPKLCPADPDMIAFIHSDIMISNIVTREERRRLTYVHNELANEEDARSA 250
DB 147 ----TGDEVDVAVSPKRG--GVASVYRDQNIYIKPVAGGAEFTALT-----TGKQALSF 194

```

```

QY 251 GVATFVLOEERDRSGYWCCKAETTSQCKILAILFEENDESEVEIHWTSPLMETRRA 310
DB 195 GVAEFIVQEBLDRFTGYWMSPEDES-----RIVYTRVDESQVDIV-----PRA 236
QY 311 D-----SFRPKGTANPKVTFKXSEIMDAERI--IDV-IDELIQPEILLEG 358
DB 237 DIGFGATVNNQRIFFRAGRPAVVDLVRDL--ASGVVALDLGANKDI----- 283
QY 359 VEYIARAGWTPPEGKYAMSILDRSOTRLQIVLISPELFIVEDDVMERORLIESVDSVT 418
DB 284 --YVARVAMSADGKTIVYVQRLSRDQKTLDLIAF-----DAAT 318
QY 419 PL--IYBETTDVININHDIFHVPQSHHEEIEFIASECTGFRHLKYTSILKESYK 476
DB 319 GAGKTLITDDPHFIEVSNFRPLTDG---FTLMGSE-KDGNQHLTRYA----- 363
QY 477 RSSGLPASPDPKCPRIKEIATSGMEVTLGSHGNSIQVDERRLVYEGKDSPLEHLL 536
DB 364 -ADGKLA-----QITGDMVPIGLEG---VDEARKVALFSASIDTPIERRL 406
QY 537 YVSVYVNPGEVTRLTLDRGYSHSCCIQHCDFISKYSNOKNPHCVSLYKLSSP-----ED 591
DB 407 YEVSYAKRQKPKALTSAGWMAAKVADNCGAFAGTSDPKTPSQGTALYSADGKRVWIEE 466
QY 592 DPTCKTEFPATILDSAGPLPDYTPPELFSPESTTGTLVGMLYKPHDLOPGKKYPTVLF 651
DB 467 NKLAEHGPYW---PYANIPQ--PEFGSKAADGETLHLEIKPFGFDPARKYPALVS 519
QY 652 IYGGPQVOIVNNRKYKVPRLNTLASLGYVVVIDNRGYSCHRGKFEKGAFTKMMQGEI 711
DB 520 YVGGPQVHQRWAKMHSSE---RFLYLAGVIFPLDRGSGNRKAKPRALDRKLGVEV 576
QY 712 DDQVEGLQYLASRYDFIDLDVYGHGMSYGYLSTLMAIMORSDFRVALAGAPVTWLFY 771
DB 577 EDQLIGAKFIASQ--PYVADADKLGVMGWSYGGFAMLMILTAEPTFKGAGAGAPTEMSLY 635
QY 772 DTGYTERYMGHPDQNGYLLGSVMAQAEKFPSPRNLLILHGFDENHFAHSTILSF 831
DB 636 DTATYTERYMGHPDQNGYLLGSVMAQAEKFPSPRNLLILHGFDENHFAHSTILSF 831
QY 832 IYRAGKPYDLOIYPOERHS 850
DB 694 LQKALILFEMAMYPGERHS 712

RESULT 3
JC5142
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
C/Species: Xanthomonas maltophilia
C/Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
C/Accession: JC5142
R:Kadashima, T.; Ito, K.; Yoshimoto, T.
J. Biochem. 120, 1111-1117, 1996
A/Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and expressio
A/Reference number: JC5142; MUID:97164011; PMID:9010758
A/Accession: JC5142
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-741 <RAB>
A/Cross-references: DDBJ:D83263; NID:g1753196; PIDN:BA11872.1; PID:g1753197
C/Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl re
C/Superfamily: dipeptidyl-peptidase IV
C/Keywords: dipeptidylpeptide hydrolase; membrane bound
F418/Domain: transmembrane #status predicted <YMW>
F610/Active site: Ser #status predicted
F685/Active site: Asp #status predicted
F717/Active site: His #status predicted

Query Match 16.0%; Score 754; DB 2; Length 741;
Best Local Similarity 28.4%; Pred. No. 1.1e-45;
Matches 214; Conservative 120; Mismatches 281; Indels 138; Gaps 26;

```

QY 133 SREELERKRIGTV-GIASYDHYQSGTFLFOAGSG--IYHVADGGQGTQOPLRPN 189
 Db 93 SDEEARRERKRIAMTGIYDWSGPDARLLFPLGGELYLDLQOEGDAVRQI----- 147
 QY 190 LVETSCPNIRMDPKLCPADPDWIAFIHSNDIWINIVTEEBRRLTYVNHLEAMBEDARS 249
 Db 148 ---THGEFATDAKISFPG-GFVSFI RGNLWIVDLASGRQWLT-----ADSGT 193
 QY 250 A--GVATFVLOEFDRYSGVYWCPCRAETTPSGKILRLIYEBNDESEV-----EIIHV 300
 Db 194 TIGNGIAEFVADDEEMDRHTGYWMAPDDSA-----IAYARIDESPVYQKREYVAD 244
 QY 301 TSPMETRRADSFRRPKGTANPKYTFKMSIEMIDAEGLIDVIDKELOPEELIFEGVE 360
 Db 245 RTDVEQ-----RYPAGDANVQVKGVI SPAEQAQTQWIDLGKEODI----- 287
 QY 361 YIARAGWTPGKYAAILDRSQTRLQIVLISPFLFIPVEDDWMEROLIESVPDVTPL 420
 Db 288 YLARWMDPQHLSTQ-RQSRDQKLDLVEVLT-----ASNOQR----- 325
 QY 421 IYVEETDIWINIHDFHVPFPOSHEEIEFI FASCKTGFRHLKYKITSILKSKYKRSQ 480
 Db 326 VLAHTSEPTWVPLNHSRLFL-----DDGSIIMSSS-RFGQHLRYI-----DSKAKA-- 372
 QY 481 GLPAPSDKCPKEEIAITSGEWEYLGRHGSNIQVEVRLVYFEGTDSPLHHLHYVS 540
 Db 373 -----ALTHGNMSV---DELLAVDEKAGLAFPAIGESABESQIYAVP 412
 QY 541 YVNGEVRTLTDRGYSHSCISQHCDFPISKYNQKNPHCVSLYKLSPPEDPTCTKEF 600
 Db 413 -LOGGQFQRLSKAPMGHSASFARNASVYVDSMNSNSTPQJLELFRANG-----EK 461
 QY 601 WATLL--DSAGPLPYT-----PELFFSESTTGF-LYGMLYKPHDLPQKKYPTVL 650
 Db 462 IATLVENDLADPKAPYARREARQREVEFGTLTAADGKTLNYSVYKPAFGPPAKRYPAV 521
 QY 651 FIYGGPOVLVNNRFGVKYFRILNT-LASLGYVVVVIDNRGSCHRGLKFEBAFKYKQI 709
 Db 522 YVYGGPSQVYTDSPFGDHLFNQYLAQGGIVVFSLNKSTPRGRGFGALYKQCTV 581
 QY 710 EIDDQVGLQYLAIRYDFIDLRVGIHMSYGYLSLMAQORSDFRVAIAGAPVTLMI 769
 Db 582 EVADQLGVAWL-KQGPVWDPAIRIGVQMSNGYITMLLAKASDVAICGAGAPVDWG 640
 QY 770 FYDGTGFRYVGHGDQNEQGYLLGSVAMQAEKFPSEPNRLLILHGFIDENYHFAHTSL 829
 Db 641 LYDSHYTERYVDLPARRDAGYREARVLTIEGLRSP--LILLHGMADDNVLFTNSTSIM 697
 QY 830 SFLVRAGKPYDLOIYPOBRHSIRVPESGEHYEL 862
 Db 698 SALQKQGFPELMTYPPAKHGLSGADALHRYRV 730

RESULT 4

X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Flavobacterium meningosepticum
 C:Species: Flavobacterium meningosepticum
 C:Date: 28-Oct-1996 #sequence, revision 13-Mar-1997 #text-change 20-Jun-2000
 C:Accession: S66261
 R:Kabaashima, T.; Ito, K.; Yoshimoto, T.
 Arch. Biochem. Biophys. 320, 123-128, 1995
 A:Title: Cloning, sequencing, and expression of the dipeptidyl peptidase IV gene from *Flavobacterium meningosepticum*
 A:Reference number: S66261; MUID:55314307; PMID:7793970
 A:Accession: S66261
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-711 <KAB>
 A:Cross-references: EMBL:DA42121; NID:9577283; PIDN:BA07702.1; PID:9577284
 C:Superfamily: dipeptidyl-peptidase IV
 C:Keywords: dipeptidyl-peptidase hydrolase

Query Match 13.2%; Score 620; DB 2; Length 711;
 Best Local Similarity 25.2%; Pred. No. 4e-36;

Matches 195; Conservative 134; Mismatches 284; Indels 160; Gaps 26;
 QY 145 IGTGIASYDYHQG-----SGTF-----LFOAGS-----GIYVXKG 176
 Db 54 IEPRTIAKSYTSQKKNYIDGSGQYTPNSDESKILLQSSQSYTHRSFLGKEVND 112
 QY 177 GPGQFTQOPLRP-NLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWINIVTEEBRRLTY 235
 Db 113 -----LKSRTVVSINNNANWIGEPKSP-DSKVAFIADNNLFYQDLNGLKTLQITL 162
 QY 236 --VNHLEAMBEDARSAGVATFVLOEFDRYSGVYWCPCRAETTPSGKILRLIYEBNDES 293
 Db 163 DKKKNEIIN-----GLGQWVYEEBFGHADYQW-----NKAGDALVFRPDERKVP 208
 QY 294 EVEI-IHTYS--PMLETRADSFRPKGTANPKYTFKMSIEMIDAEGLIDVIDKELO 350
 Db 209 EINIPTIYQNIYPKLMT-----YKTPKKGEEESATAYLYQI--SSGK----- 249
 QY 351 PFEILFEGVE--YIARAGWTPGKYAAILDRSQTRLQIVLISPFLFIPVEDDWMERQ 408
 Db 250 SAQINFGSGSEKYYIPQLPQTWANDIYVATANRHQNKVDLKNRK----- 295
 QY 409 LIESVPDVTPLIYEETDIWINIHDFHVPFPOSHEEIEFI FASCKTGFRHLKYKITS 468
 Db 296 -----TAAVSKLFTETDNAMIEITDNLTFMF-----LDDNSFLWASE-RDGRHLHYDA 343
 QY 469 ILKSKYKRSQGLPAPSDKCPKEEIAITSGEWEYLGRHGSNIQVEVRLVYFEGTK 528
 Db 344 AGKLKK-----QVSKGMEIINYYGNPKTE---VITQITE 377
 QY 529 DSPLEHLHYVS--YVNGEVRTLTDRGYSHSCISQHCDFPISKYNQKNPHCVSLYK 586
 Db 378 KGSINK--VYSKININTGKTQLLSNABGNNSAASFCKTFNFINTSSYAKYPTKILDA 434
 QY 587 SSPE-----DDPTCKTKEFWATILSAGPLPYTPELISFESTTGFYLYGMLYKPHD 639
 Db 436 NGKQVKELONNDDLNLKLS-----DNFIAKEFTTIPPAADQNMAMWIKPKN 482
 QY 640 LQPGKYPVTVFIYGGPOVLVNNRFGVKYFRILNTLASLGYVVVVIDNRGSCHRGLKFE 699
 Db 483 FDPAPKTFVPMFOYSGSGSQOVANSWDGNGIMEDMLAQKGLVVCVGRGFGYKYY 542
 QY 700 GAFYKKGQIEIDDQVGLQYLAIRYDFIDLRVGIHMSYGYLSLMAQORSDFRVAIAG 759
 Db 543 KVTYKNLGKYEIEDQITAAKVLGNQ-SYVDSRIRIGIPMSYGGYNASLAKMGADVFMG 601
 QY 760 IAGAPVTLMTFYDGTGFRYVGHGDQNEQGYLLGSVAMQAEKFPSEPNRLLILHGFIDEN 819
 Db 602 IAAVAVTNMRFYDSIYTERFLQTPQENKDGDLNLSPTTYAKTL--KGFILLHGTADDN 658
 QY 820 VHFAPTSILLSFLVRAGKPYDLOIYPOBRHSIRVPESGEHYELHLYLOENL 872
 Db 659 VHFQNSMFESEBALIQNKQKQDFPMAYIPDKNHSIIGANTRPQYKQNTYILENL 711

RESULT 5

CDH26
 dipeptidyl-peptidase IV (EC 3.4.14.5) - human
 N:Alternate names: cell surface glycoprotein CD26; thymocyte-activating molecule (THAM)
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence, revision 23-Aug-1996 #text-change 18-Jun-1999
 C:Accession: S24313; B42408; B42408; B61136; S59510; I56154; S59857; S15520
 R:Kusunagi, Y.; Hayashi, Y.; Arakawa, F.; Ikehara, Y.
 Biochim. Biophys. Acta 1131, 333-336, 1992
 A:Title: Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a serine protease
 A:Reference number: S24313; MUID:92329551; PMID:1352704

A:Molecule type: mRNA
 A:Residues: 1-6,17,8-766 <MIS>
 A:Cross-references: EMBL:X60708; NID:935335; PIDN:CAA3118.1; PID:935336
 J:Darmon, D.; Lacasa, M.; Batteault, L.; Marguet, D.; Sapin, C.; Trocote, P.; Barbé, A.
 J. Biol. Chem. 267, 4824-4833, 1992
 A:Title: Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer

IV mRNA levels during cell differentiation.

A:Reference number: A42408; MUID:92165847; PMID:1347043

A:Accession: B42408

A:Molecule type: mRNA

A:Residues: 1-5, 'R', '7-436', 'S', '438-556', 'I', '558-662', 'E', '664-766' <DAR1>

A:Cross-references: GB:M80536; NID:g181569; PIDN:AAA52308.1; PID:g181570

A:Experimental source: intestine

A>Note: this sequence corresponds with the author's translation

A:Accession: A42408

A:Molecule type: mRNA

A:Residues: 1-5, 'R', '7-436', 'S', '438-556', 'I', '558-662', 'E', '664-711', 'G', '713-766' <DAR2>

A:Cross-references: GB:M80536; NID:g181569

A>Note: sequence extracted from NCBI backbone (NCIN:83986, NCBI:P.83988); this sequence

R:Correl, J.P.; Ferrero, A.; Chandraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.

Gastroenterology 101, 618-625, 1991

A:Title: Expression of sucrose-isomaltase and dipeptidylpeptidase IV in human small intestine

A:Reference number: A61136; MUID:91317403; PMID:1677636

A:Accession: B61136

A:Molecule type: protein

A:Residues: 1-15, 'X', '17-22' <GOR>

R:Boehm, S.K.; Gum Jr., J.R.; Erickson, R.H.; Hicks, J.W.; Kim, Y.S.

Biochem. J. 311, 835-843, 1995

A:Title: Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a

A:Reference number: S59510; MUID:96067599; PMID:7487939

A:Accession: S59510

A:Molecule type: DNA

A:Residues: 1-31 <EOB>

A:Cross-references: GB:S79876; NID:g195574; PIDN:AA835614.1; PID:g195575

R:Tanaka, T.; Camerini, D.; Seed, B.; Torimoto, Y.; Dang, N.H.; Kameoka, J.; Dahlberg, H.

Immunol. 149, 481-486, 1992

A:Title: Cloning and functional expression of the T cell activation antigen CD26.

A:Reference number: I56154; MUID:92325476; PMID:1352530

A:Accession: I56154

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-436, 'S', '438-766' <TRAN>

A:Cross-references: GB:M74777; NID:g180082; PIDN:AAA51943.1; PID:g180083

R:Abbot, C.A.; Baker, E.; Sutherland, G.R.; McCaughan, G.W.

Immunogenetics 40, 331-338, 1994

A:Title: Genomic organization, exact localization, and tissue expression of the human CD

A:Reference number: S59857; MUID:95012454; PMID:7927537

A:Accession: S59857

A:Molecule type: DNA

A:Residues: 1-436, 'S', '438-766' <ABB>

A:Cross-references: EMBL:U13734

A:Gene: GDB:DPP4

A:Cross-references: GDB:125239; OMIM:102720

A:Map position: 2q24.3-2q24.3

A:Introns: 2/3; 32/1; 65/1; 95/3; 122/3; 140/2; 164/3; 205/1; 258/3; 296/2; 341/3; 356/3

C:Superfamily: dipeptidyl-peptidase IV

C:Keywords: dipeptidylpeptide hydrolase; glycoprotein; homodimer; proteinaase; transmembr

F1-6/Domain: intracellular #status predicted <INT>

F12-28/Domain: transmembrane #status predicted <TMN>

F129-766/Domain: extracellular #status predicted <EXT>

F185,92,150,219,229,281,321,520,685/Binding site: carbohydrate (asn) (covalent) #status

F1630,708,740/Active site: Ser, Asp, His #status predicted

Query Match

Best Local Similarity 11.3%; Score 529; DB 1; Length 766;

Matches 186; Conservative 109; Mismatches 280; Indels 128; Gaps 32;

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy 373 YAMSIILDRSQRLQVILISPELFIPEVDVMEORQIESVPSVTLPIEETTDITWN 432

Db 307 -----TORISIQWL-----RIQNY--SYNDICDYDESSGRM-N 338

Qy 433 HHDIFVHPQSHSEELIFASECKTGFRH-----LYKITSILKESKYKSSGGLPAP 485

Db 338 C-----LVARQRIEMSTTGWVGRPRSEPHFLDNGSPFKIIS--NEEGYRIH----- 384

Qy 486 SDFKPCPIKEELITGSEVVLGRHNSIQVDEVRLLVPEGR-KDSPLEHLLVSVYVNP 544

Db 385 CYFOIDKDCDTRITGQWMEVIG-----IEALTSVLYYISNEYKMGPGGRNLYKIQLIDY 439

Qy 545 GEVTRLTGRVGHSCIS-QHCDPFRSKSNQKNH---C-----SVLYLSSPEDPTCK 596

Db 440 TKTTL-----SEIIMPERQVYVSFSKAKYQLRCSGPGPLPYLTHSSVNDKGR 492

Qy 597 TKEFWATILDSAGPLPDYTPPE---IFSESTTGFTLYGMLYKPHDLQCKKYPTVLFY 653

Db 493 VLED-NSALDKM--LQNVQMPSEKLDFTILNETKF-WYQMLIPH-FDKSKYPLLDVY 547

Qy 654 GGPQVQVNNRKKYKFRIN---TLASGLYVVV-IDNRGSHRGLKFBGAFKYMGOI 709

Db 548 AGPCSQ-----KADTVRLNATYLASTENITVASPDGRSGYQGDKIMHAINRRLGTF 601

Qy 710 EIDDOVEGLQYLASHRYDIDLDVGIHGMVSGYGLSLMALMQRSDFRVAIAAPVTLMI 769

Db 602 EVEDDIEAARQ-SMGGVVDNRILIMKMSYGYTTSVNLGSGSGVFKGCIIVAPSRKE 660

Qy 770 FYDTGYTRYWG--HPDNEQGYLYGSVAMAQKEPSEBNRLLLHGFLENVHFAHPSI 827

Db 661 YDSVYTERYWGFLPPEEDLDHYRSTVMSRAENF--KQVEYLLHGTADNVHQQSQSQ 718

Qy 828 LLSFLVRAGKYVDLQYQERHSTRVPSGEHYEHLHLYOE 870

Db 719 ISKALVDVGVDPQAMWYDEDHGIASTAHQHIYTHMSHFITQ 761

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

373 YAMSTLL---DRSQRLOIVLISPEL--FIPVEDVWER-----ORLIESPDSVTPLELI 422
 Db 327 ---SVLVFTNRNSSTICFARLLDRLTSLHITKTECLEGEWYEQSAMP--LNNSLV 381
 Qy 423 YEETDVIWINDIHFVFPQSHHEIEFIPASECTGRRHLKYLSILKESYKSSGGL 482
 Db 382 WENMSD---GYFDIALADYNNLAFIP-----NASS 410
 Qy 483 PAPSDKCPKEELIAITSGEWEVLGRHSNIOVDEVRRLVVEGTRKSPLEHLLVYVYV 542
 Db 411 P-----ILYTSGAMDVT---DGIHIDSDPGNVYFLAILKOSTERHLYVVS-L 454
 Qy 543 NPGVTRLTDRGY---HSCCISQCHDFIKSYNQNPKHCVSLYKLSPEDDPTCKTKE 599
 Db 455 DTLEIYGLTDGDEGEYVSTSPRGDFVLYNHGPDPV---MQLRSTXDKOYCSLE 510
 Qy 600 FMATILDSAGPLPDYTPPEI---FSPESTTGFTLYGMLYKPHDQPKKXPTVLFITGG 655
 Db 511 TNSRLKQO---LSSITLPSVEYCKLTENDTT-FNF--MERRPRNPDVNRKTPVLEFAYG 564
 Qy 656 PVOVLVNNRFKGVKYPRLNTLASLG-----YVVVVIDNRGSGHGLKFEAGAFYKMGQI 709
 Db 565 PGSQOV-----AKLFRVDPQAVLASHDPDEFIVTLDRGTGFNGNAFRYSVRHLGEW 618
 Qy 710 EIDDOVEGLQVLAARVPIDLDRYGHNQSGYGLSLMALMQRSDIPRVAIAGAPVTIM 769
 Db 619 ESYDQOQKGFMAW-LRPVDENHWGIMWGSYGYLTLKTL-ETQDVFSYGAAVAPVIMWR 676
 Qy 770 FYDGYTERVYGHDPDQEGYVYLSVAMQAEKFPSEPRNLLHGFEDENHFAHTSILL 829
 Db 677 LYDSVTERVYMDLPQYNKEG-YKNSQIHDIYEF-KOLKRFVAFHGTGDVNHFGSHMIM 734
 Qy 830 SFLLVRAG-KPYDLQIYQERHSIRVPSGEHYELHLHYLOENLGSRIAL 879
 Db 735 DGLMLANCYNDMAVFPDSASHI---SYNNASLSIYHRLSEWIGDALGRI 781

RESULT 7
 S23752
 dipeptidyl-peptidase IV (EC 3.4.14.5) alpha chain - mouse
 N/Alternate names: CD26 alpha subunit, THAM alpha subunit
 C/Species: Mus musculus (house mouse)
 C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999
 R/Accession: S23752; A46465; A56030
 R/Marguet, D.; Bernard, A.M.; Vivier, I.; Darmon, D.; Naquet, P.; Pierres, M.
 J. Biol. Chem. 267, 2200-2208, 1992
 A/Title: cDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto-di-
 A/Reference number: S23752; MUID:92129288; PMID:1370813
 A/Accession: S23752
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-760 <MAR>
 A/Cross-references: EMBL:X58384
 R/Vivier, I.; Marguet, D.; Naquet, P.; Bonicel, J.; Black, D.; Li, C.X.; Bernard, A.M.;
 J. Immunol. 147, 447-454, 1991
 A/Title: Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptidase
 A/Reference number: A46465; MUID:91302787; PMID:1712807
 A/Accession: A46465
 A/Status: preliminary
 A/Molecule type: protein
 A/Residues: 1-20 <VIY>
 A/Experimental source: M14.T thymoma cells, Swiss nu/nu
 R/Bernard, A.M.; Mattei, M.G.; Pierres, M.; Marguet, D.
 Biochemistry 33, 15204-15214, 1994
 A/Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.
 A/Reference number: A56030; MUID:95092780; PMID:7999781
 A/Accession: A56030
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 746-760 <BER>
 A/Cross-references: GB:U12620

C/Genetics:
 A/Gene: CD26
 C/Superfamily: dipeptidyl-peptidase IV
 C/Keywords: dipeptidyl-peptidase hydrolase; glycoprotein; transmembrane protein
 F.213,223,315,514,679/Binding site: carboxylate (asn) (covalent) #status predicted
 F.624,702,734/Active site: Ser, Asp, His #status predicted

Query Match 11.0%; Score 518.5; DB 1; Length 760;
 Best Local Similarity 25.0%; Pred. No. 8e-29;
 Matches 203; Conservative 107; Mismatches 306; Indels 197; Gaps 37;

131 MYSEELRERKRIIGVGI-----ASYDHQSGTFLF-----OAGS 168
 Db 67 LYKQENMLLNAAHGNSSIFLENSTFSPGYHSVDPRLVLELVYVQMRHSYTSY 126
 Qy 169 GIYVHKGQPGGFQOPLRPLVLTSCPNIMDRKLCAPADWI-----AFHSND 219
 Db 127 NIYVNR-----RQLTEKIPN-----NFWITWSPGKHLAYVND 165
 Qy 220 IWINIVTREBRRLTYVHNELANMEEDARSAGVATFVLQEE-FPRYSGVWNCPK----- 272
 Db 166 IYVAVEPHLPBHRIT-----STGEENVITNGITDWVVEEYFAGYSLMMSPNNTFLAY 219
 Qy 273 AETTPSGEKILRIIYEENDESEVEIIVTSPMLETRRADSRFPYKGTANPKYTPKSGEI 332
 Db 220 AQFNDTGVPILTEYSF-YSDS-----LQYPKTWMIYPYKAGANPTVKFFI--V 265
 Qy 333 MIDAEGRHIDVDEKLQPFELFEGVEYIARAWTPBGRKA-----WSILLDSQ 383
 Db 266 NIDLSSSSSAAPQIAPAPASA-RGDHYLQDVVAATEERISLOWRRIQNYSWAICDY 324
 Qy 384 TRLOIVLISPELFIPEDEDVWERQRLIESVDSVTPILITYETDVIWINDIHFVFPQS 443
 Db 325 DKIMLVNCCS-----EQQH-----EMSTGWVG--RFRPAEYHF 358
 Qy 444 HEEIEFTFASCTGTRHLYKITSILKESYKSSGGLPAPSDPK-CPIKEIAITSGE 502
 Db 359 TSDGSFYKLTISDQGYKHCHF-----PKDKKQCTF-----ITGGA 395
 Qy 503 WEVYGRHSNIOVDEVARLYVFBGT-KDSPLEHLLVYVSNPBEVTRLTGRGSHSCCI 561
 Db 396 WEVY-----STELTSDLYLISNQYKEMPGGRKLYKI-----QLTQ--HTNVKCL 439
 Qy 562 S-----CHCDPFIKSYNQNKNP-----C-----VSLYKLSPEDDPTCKTKEFMATILDSAG 609
 Db 440 SCDLNPRCQYVAVSPFKAKYQGLGQMGRLPYTLIHRSDHKELVLE-----DNSA 493
 Qy 610 ---PLPYTPPE--ITSFSTTGFTLYGMLYKPHDQPKKYPVLFYGGPOVOLVNN 663
 Db 494 LDRWLDQVWPSKLDPIVNETRF-WYQMLIPPH-FDKSKKYPDLLDVVAGPSCQ----- 547
 Qy 664 RFGVKKYFRLN---TLASLGVVVV-IDNNGSGHGLKFEAGAFYKMGQIHIDQVGLQ 719
 Db 548 --KADASFRLNMTYLASTENITVASFDGSGSGQKIMAINRRLGTELEVEDQIEAR 605
 Qy 720 YLASRYDFIDLDVGVHGMVSGYGLSLMALMQRSDIPRVAIAGAPVTIMFYDGYTERY 779
 Db 606 QAV-KMGFVDSKVVAVLMGWSYGVYVSMVLGSSGVYKCGIIVAPVSMVEYDSVTERY 664
 Qy 780 MG--HEDQEGYVYLSVAMQAEKFPSEPRNLLHGFEDENHFAHTSILSLFVRAGK 837
 Db 665 MGLPIEDNIDHNRNSTVMSRAHF--KQVEYLLIHGTADNVHFGQSAISKVLVDAGV 722
 Qy 838 FYDLQIYQERHSIRVPSGEHYELHLHYLOE 870
 Db 723 DFGAMVYTBEDHGIASSTAHQIYSHMSHPLQ 755

RESULT 8
 A39914
 dipeptidyl-peptidase IV (EC 3.4.14.5), membrane-bound form precursor - rat
 N/Alternate names: GP110, membrane glycoprotein 110K; OK-61
 N/contains: dipeptidyl-peptidase IV, soluble form

A:Experimental source: clone T23F1

C:Genetics:

A:Gene: CESP:T23F1.7a

A:Map position: 5

A:Introns: 13/3; 52/1; 111/3; 151/2; 177/3; 346/3; 504/1; 537/1; 628/2; 683/1; 723/1

C:Superfamily: dipeptidyl-peptidase IV

Query Match 9.5%; Score 445.5; DB 2; Length 779;

Best Local Similarity 23.0%; Pred. No. 1.4e-23;

Matches 193; Conservative 126; Mismatches 292; Indels 229; Gaps 38;

```

QY 78 PHSRIYLAAMS-----GENRENTLFYSEIPKTIINRAVLM,SMVKFLDLFOATLDYMY 132
D 116 PSARRYVAMMDHAPNPGMNPQNETFHLKTYNNR-----ITYDGL- 158
QY 133 SREBBLERKRIGTVAGSYDYGSGTFLPQAGSGIYHKDGGPGFTOQPLRPLVE 192
D 159 -RKESV-----IQAFKMGKFNDFVFNKIKIY----- 187
QY 193 TSCNINRMDPLCPADPMIAPFHSNDIWSINIVRERRLTYVHNEIANEEDARSAGV 252
D 188 QSSP-----EEGGLTRVSNNGEHTVD-----GL 210
QY 253 ATFLVLOE-FDRYSGYWMCPKAETTPSGKILRLIYEENDESEVEIIVTSPMLETRAD 311
D 211 FDMYEEBIFGRKXAMWMTK-----GDQAYASYDWHLTRKWSL-KTYHRLPEPYID 262
QY 312 -SEFYPT-----GTANPKVTFKSEIMIDABGRIDVIDKELLQPEFLL-----FEGV 359
D 263 TNFHYPIKFAVLPTLYTILSNKKE-----QSRQLDQOLDSLYHYLLAVKMLEINGT 317
QY 360 EYIRAGWTPGKAFAMSLI-----DRSQTRLQIVLISPELFIPVEDVMERQVLIBVP 414
D 318 EQLVSV-WT-NRQNEVALTICMDIATCRLEFE----- 349
QY 415 DSVPLIYEETDIIWIMIHDIHFVFPQSHHEIEFIIPASECKTGFRHLYKITSILKSK 474
D 350 -----YKYASKRWLT-HDDFHSI-TSFEDTLFLLP-----HDKXDNMFOQVAS 391
QY 475 YKRSSGGLPAPSDKCPKEIKIATISGMEVLAGRHGNSIQVDEVRLVYFEGTQDSPLH 534
D 392 LRLSHGQLRTEK-----PLNLGHVDTVSINGINKET-----RTIFFHAAPKPSHR 437
QY 535 HLIVVSYVNPGEVTRLDRGYSHSCIS-----OHCDPFIKSKXNKNPCVSLYKLSPED 591
D 438 SLFSYS-----LADSRNSAYCISCSLKNCITMAQOMDDMKATAYSCCKPAAHT 488
QY 592 -----DPTCKTK-----EFWATILDSAGPLPYTPPEIFSFESTTGFTLYGMLYK 636
D 489 AIVNLTMDSDKTEHANMLYDKTYQNRVEAG-LPVIKETI---KISDFDALIKLSI 544
QY 637 PHLDQPGKY---PTVLFYIGSPQVQVNNRFKGVKTYFRLLNTLASLGVVVV-IDNRGSC 692
D 545 PKDIYNDRKQALPLIYHVGSGPNDQ---NTEKATQIGIEVVASASQALILRIDGSGG 601
QY 693 HRGLKFGAFKYGKQGLEIDQVGLQYLASRY-DFIDLDRVGIHSGSYGYLSLMLMQ 751
D 602 GRGKYSATYIGQGLYVEVEQIKAIKIVLRLYHLLDARVAVFGMSYGGFMTLSWNE 661
QY 752 RSD-IFRVALIAGAVTLMIIFYDTGYTERYMGHPDQNGYLLGSVAMQAEKFPSEPNRL 810
D 662 APEQFFKCAVSAVAVTFAYAYDATYTERYMG--DAPLESY--SPVTKKLDNFKS--TRL 715
QY 811 LLLHGLFDENVHFAHTSILSLFLVRAAGPYDQIYPOERHSIRVPSGEHYELHLLHYLQ 870
D 716 LMGHGLDDNVHFOHSALILDELONRGVDFLMLVYPNOAHSLSRTS--HVVGKMTFLRQ 773

```

RESULT 13

A41793

dipeptidyl aminopeptidase-like protein (EC 3.4.14.-) short form, DPPX-S - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C:Accession: A41793

R:Wada, K.; Yokotani, N.; Hunter, C.; Doi, K.; Wenthold, R.J.; Shimaaki, S.

Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992

A:Title: Differential expression of two distinct forms of mRNA encoding members of a di-

A:Reference number: A41793; MUID:92108018; PMID:1729689

A:Accession: A41793

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-803 <MAD>

A:Cross-references: GB:M76429; NID:9408719; PIDN:AAC41623.1; PID:9408720

A:Note: sequence extracted from NCBI backbone (NCBI:75138)

C:Superfamily: dipeptidyl-peptidase IV

C:Keywords: dipeptidylpeptide hydrolase; glycoprotein

F:257/342/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.4%; Score 440; DB 2; Length 803;

Best Local Similarity 24.4%; Pred. No. 3.6e-23;

Matches 176; Conservative 114; Mismatches 317; Indels 114; Gaps 29;

```

QY 171 YHVKDGGPGFTQQLRPNLVETSCPNIRMDPKCPADPMIAPFHSNDIWSINIVTRE 230
D 156 YVVLSTKIPHGDPQSLDPEVSNALKQYAGMGPK-----GOOLIFENNIIYCAHVGRQA 210
QY 231 RRLTYVHNEIANEEDARSAGVATFLVLOE-FDRYSGYWMCPKAETTPSGKILRLIYEE 289
D 211 IRV-----YSTGEVITYNGLSDMLYEEILKTHIAWMSPDG-----TRLAYAT 255
QY 290 NDESEVELIHV---TSPMLETRADSPRYPTGTANPKVTFKSEIMIDABGRIDVIDK 346
D 256 INDSRVPMELPTTGSGYPT-AKPHYHPRAGGENPISLH---VIGLNGPHTDL--- 306
QY 347 ELIOPPELPEGEVYIARAGTPEKTA-WSILDSQTRLQIVLISPELFIPVEDVM 404
D 307 EMTPDDPERMEX-YITVWKATSTKYAVNV---LSBAQN-----VS----- 344
QY 405 ERQLISVDPSPVPLII---YEETDIIWIMIHDIHFVFPQSHHEIEFIIPASECKTGRH 462
D 345 -----LTLCDATGTCCKHDESEAML-----HRQNEBVFVS---KDQ-RK 383
QY 463 LYKITSILK-ESKYKSSGGLPAPSDKCPKEIKIATISGMEVLAGRHGNSIQVDEVRL 520
D 384 FFFRAIPQGGQGFYHITVSSSQPNSSNDIQ---SITSGDMVY---KILSYDEKRS 436
QY 521 LVYEGTKDSPLEHLLVSVNPEVTR-----LDRGYSHSCISQHCDFPIKSKXN 574
D 437 QIYFLSTEDLPERRQLYSASTV-GSFNRQCLSDLVNCTYFSAFSPGADFFLLKCEG 494
QY 575 QKNHCYSLYKLSPEDDPTCKTKE-FWATILDSAGPLPYTPPEIFSFESTTGFTLYGM 633
D 495 PGVP-TVSVHNTDKKMPDLEINEHQAISDKQMPKVEYRKLE-----TDDYNLPDQ 547
QY 634 LYKEDLQPGKYPTVLFYIGSPQVQVNNRFKGVKTYFRLLNTLASLGVVVV-IDNRGSC 693
D 548 ILKATPTDPAHYVLLVVGTPGSGVAKEF-AVMEWTVMVSHGAVVVCGRSGGF 605
QY 694 RGLKFSAPFKYKQGLEIDQVGLQYLASRYDFIDLVYGIHSGSYGYLSLMLMQRS 753
D 606 QGTRLLEHVRRLRGLSELEKQMEAVRYVL-KEPYIDKTRVAVFEGKDYGYLSTYLLPAKG 664
QY 754 D-----IFRVALIAGAVTLMIIFYDTGYTERYMGHPDQNGYLLGSVAMQAEKFPSEPNRL 809
D 665 DGAQPVVSCGSAISPTIDFKLYIASERLYGLHGLNRAVEAKVHRSAL-EGQQF 722
QY 810 LLLHGLFDENVHFAHTSILSLFLVRAAGPYDQIYPOERHSIRVPSGEHYELHLLHYLQ 869
D 723 LVTHATDEKIHQHTAELLTQILKGRANSLQIYPPESHYFSSAALQCHLHRSILGFV 782
QY 870 E 870
D 783 E 783

```

RESULT 14

168600
dipeptidyl aminopeptidase like protein - human
C|Species: Homo sapiens (man)
C|Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
C|Accession: 168600
R|Yokotani, N.; Doi, K.; Wenthold, R.J.; Wada, K.
M|Title: Mol. Genet. 2, 1037-1039, 1993
A|Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-related
A|Reference number: I54331; MUID:93372805; PMID:8103397
A|Accession: 168600
A|Status: preliminary; translated from GB/EMBL/DBJ
A|Molecule type: mRNA
A|Residues: 1-803 <RES>
A|Cross-references: GS:M96860; NID:g306707; PIDN:AAA35761.1; PID:g306708
Superfamily: dipeptidyl-peptidase IV

Search completed: October 15, 2003, 17:14:45
Job time : 31.176 secs

0

0

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 16:57:30 ; Search time 13.9931 Seconds
(without alignments)
2964.152 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700

Sequence: 1 MAAAMETBOLGVETFTADG.....HLHYIQENIGSRNALAKVI 882

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127663 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127663

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	569.5	12.1	765	1	Q9N217 felis silve
2	566	12.0	765	1	P81425 bos taurus
3	525	11.2	766	1	P27487 homo sapien
4	523.5	11.1	767	1	DPPE_RAT
5	519.5	11.1	793	1	YEA8_SCHPO
6	517.5	11.0	760	1	DPPE_MOUSE
7	484.5	10.3	818	1	DAP2_YEAST
8	478.5	10.2	761	1	SEPR_MOUSE
9	465	9.9	931	1	ST13_YEAST
10	454	9.7	760	1	SEPR_HUMAN
11	440	9.4	863	1	DPPE_BOVIN
12	435	9.3	804	1	DPPE_MOUSE
13	431	9.2	859	1	DPPE_RAT
14	427	9.1	865	1	DPPE_HUMAN
15	293	6.2	853	1	YDZF_SCHPO
16	249.5	5.3	657	1	YUML_BACSU
17	172	3.7	732	1	ACPH_RAT
18	171	3.6	732	1	ACPH_RAT
19	161	3.4	732	1	ACPH_PIG
20	150	3.2	740	1	YL31_CAEEL
21	139	3.0	689	1	PPCE_AERHY
22	134	2.9	722	1	Y174_RICPR
23	134	2.9	1014	1	BGAL_BACHD
24	125.5	2.7	436	1	PAPA_CAVPO
25	124.5	2.6	1835	1	DURI_YEAST
26	123.5	2.6	705	1	PPCF_FLAME
27	123	2.6	1702	1	DPOI_THELI
28	122	2.6	705	1	Y4S0_RHISN
29	119.5	2.5	1019	1	CA26_HUMAN
30	118.5	2.5	705	1	PPCE_FLAME
31	118.5	2.5	710	1	PPCE_HUMAN
32	118.5	2.5	1436	1	DPO3_STEAM
33	118.5	2.5	1438	1	DPO3_STEAM

34	118	2.5	1285	1	TOXA_PASMU
35	115	2.4	1047	1	POL_STVAL
36	115	2.4	3144	1	VP13_YEAST
37	114.5	2.4	710	1	PPCE_MOUSE
38	113.5	2.4	2483	1	COA2_HUMAN
39	112.5	2.4	835	1	NLG2_HUMAN
40	111.5	2.4	465	1	EGLC_RHIME
41	110.5	2.4	836	1	NLG2_RAT
42	110.5	2.4	4349	1	FAT2_HUMAN
43	110	2.3	1295	1	BXAI_CLOBO
44	110	2.3	1861	1	MAP2_RAT
45	110	2.3	2410	1	MOXI_SCHPO

ALIGNMENTS

RESULT 1
ID DPPE_FELCA STANDARD; PRT; 765 AA.
AC Q9N217;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26).
GN DPPE OR CD26.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Periphera blood;
RX MEDLINE=20094000; PubMed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
RA Mikami T., Takahashi E.;
RT "Molecular cloning and sequencing of a cDNA encoding the feline T-cell
RT activation antigen CD26 homologue.";
RL Immunogenetics 50:366-368 (1999).
CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline (By similarity).
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-
CC xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form (By similarity).
CC -1- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC by proteolytic processing (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB023952; BAA92344.1; -
CC MEROPS; S09.003; -
CC INTERPRO: IPR002469; DPPV N term.
CC INTERPRO: IPR001375; Peptidase S9.
CC INTERPRO: IPR002471; Pro_endopep_ser.
CC DR InterPro: IPR000379; Ser_estrs_site.
CC DR Pfam: PF00930; DPPV N term; 1.
CC DR Pfam: PF00326; Peptidase S9; 1.
CC DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
CC DR Hydrolyse; Aminopeptidase; Dipeptidase; Serine protease;
CC Transmembrane; Glycoprotein; Signal-anchor.
CC CHAIN 1 765
CC DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM

```

FT CHAIN 38 765 (MDP).
FT DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT (SDP) (BY SIMILARITY).
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 7 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 30 765 EXTRACELLULAR (POTENTIAL).
FT ACT SITE 629 765 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 707 707 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 739 739 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 765 AA; 88213 MW; 3EFCB98A22B175D9 CRC64;

Query Match 12.1%; Score 569.5; DB 1; Length 765;
Seet Local Similarity 23.7%; Pred. No. 8,2e-34;
Matches 225; Conservative 114; Mismatches 316; Indels 221; Gaps 42;

71 KRNDPDGPHSDRIYVLAAMSGENRENTLFYSEIPKTNINRAAVMLSMKPLDLFOATLDYG 130
30 KGNDA-AADSRRTYTLDTYKNTFRVXYS-----LNM-----VSDHDY- 67
131 MYSREELLRRKRIGTVGI-----ASYDHOGS-----GTFL-----F 164
68 LYKODNNILFNAEYGNSSIFLENSTPDEFHSINDVSPPGQFILLEVNVKQMRASY 127
165 QAGSGIYHVKOGPGFTGOQPRPLVATSCNTRMDKLCRADPDW-----AFI 215
128 TASYDIYDLNK-----RQITTEKIPN-----NQWTWSPBEGHKLAVY 166
166 HSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLOEE-PRYSGYWMCPRAE 274
167 WKNDVYVKNBPNSSHRTWTG-----EENALYNGIADWYEEIIFAYISALMWSPGCT 220
275 TTPSGGKILRIIYENDESEVEIIT--VTSBMLETRADSRIRPKXTANPKYTFKXSE 331
221 F-----LAYAFNDTOYVPLIEFSYSDESILOYPTMTIRPYKAGANAP--TVKLFV 269
332 IMIDAEGRIIDVIDKEILIQFEIILPEGVEYIARAGCTEGKYAMSILDRQOTRLQIVLI 391
270 IKTDNLNNTNATSVETTPPAAML-TGDIYLCQDVTWANEKRS----- 311
392 SPELFIVEDDVMERQRIIESVPSVTPLIYEETTDIMINIHDFHVPQSHHELEIFI 451
312 -----LQWLRIQNV--SYMDIRDYNNSGTGKIS-----SAAQHILM- 347
452 PASCKTGF-----RHLYKITSILKESKYKXSSGGLPAPSPFKPIKEEI 496
348 ----STGQWGRFPRAEPHFTSDGKNFYKIS--NEDGYKHI-----CRFOIDKDKCCT 394
497 AITSGEWEVLGRHSGNIQVDEVRLVYPEGT-KDSPLFHLHYVVSYYVNGEVTRLTDQGY 555
395 FITGAWMEVIG-----IEALTDTYIYISNEYKMGPGRNLYKI-----QLND--Y 438
556 SHSCCIS-----QHCDPFIKSYNOKNPH--C-----VSLYKTSPPEDDPTCKTKEFWAT 603
439 TKVACTSCELKPECCQYVSFSKAKYQURCGPGPLTYLTRSSNDEELRYLED-NS 497
604 ILDAGPLPDYTPPE---IFSFESTGTFTLGMLYKPHDLPKGYKPYVLFYGGPQVQL 660
498 ALDKM--LOEVQMSKLDFTIILNFKF-WYQMIPLRP-FPTSKYPLLDIVYAGPQSG- 552
661 VNNRFGYKYEPLN---TLASLGYVVVV-IDNRGSGHGLFEGAFYKFMQOILDDOVE 716

```

```

Db 553 -----KDAIFRLNWTATYLASTENIIVASFDRGSGYGGDKIMHVAVNRRLGTFEVEDQIE 607
Qy 717 GLQYIASYDPELIDDRVGHGMSYGYLSUAMALWQSPIDFRAVATAGAVTLMIFPDQYVT 776
Db 608 AARPD-SKMGVDDRRKRLAIWGMSTGYGVTSVNLGAGSSVFQGLAVAFVSMEYEDSVYT 666
Qy 777 ERYMGHP-DONEGQYVLSGVAM--QAEKPPSPNRLILHGLFENDEVHFAHSILLSFLVR 834
Db 667 ERYMGLPFPQNDLNYKNSYMSRAENF--KQVEYLHIGTRADNVHFGQSAQISKALVD 724
Qy 835 AKRPDIQIYPOERHSIRVPESGEHYELHLHYLOE 870
Db 725 AGVDFOAMWYDDEHGIASGPAHQHIYTHMSHFQKQ 760

RESULT 2
DPP4_BOVIN
ID DPP4_BOVIN STANDARD; PRT; 765 AA.
AC B81425; Q8MWG8;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26) (Adenosine deaminase complexing protein) (ADCP-I)
DE (Activation molecule 3) (ACT3) (WC10).
GN DPP4 OR CD26.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphocytes;
RX MEDLINE=22067734; PubMed=12073152;
RA Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J.,
RA Bohach G.A.;
RT "Molecular characterization of bovine CD26 upregulated by a
RT staphylococcal superantigen.";
RN Immunogenetics 54:216-220(2002).
RL [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
RC TISSUE=Thymus;
RX MEDLINE=22021197; PubMed=11981836;
RA Gliddon D.R., Howard C.J.;
RT "CD26 is expressed on a restricted subpopulation of dendritic cells in
RT vivo.";
RL Eur. J. Immunol. 32:1472-1481(2002).
RN [3]
RP SEQUENCE OF 1-24.
RC TISSUE=T-cell;
RX MEDLINE=21482004; PubMed=11598101;
RA Lee S.-U., Ferens W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K.,
RA Naessens J., Bohach G.A.;
RT "Identity of activation molecule 3 on superantigen-stimulated bovine
RT cells is CD26.";
RL Infect. Immun. 69:7190-7193(2001).
RN [4]
RP SEQUENCE OF 537-546.
RC TISSUE=Kidney;
RX MEDLINE=96293306; PubMed=9629661;
RA Ben-Shooshan I., Parola A.H.;
RT "The CP-I subunit of adenosine deaminase complexing protein from calf
RT kidney is identical to human, mouse, and rat dipeptidyl peptidase
RT IV.";
RN Comp. Biochem. Physiol. 119B:289-297(1998).
RP FUNCTION: Removes N-terminal dipeptides sequentially from
RP polypeptides having unsubstituted N-termini provided that the
RP penultimate residue is proline. Binds and regulates the activity
RP of ADA.
CC -!- CATALYTIC ACTIVITY. Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC xcc, from a polypeptide, preferentially when xbb is Pro, provided
CC xcc is neither Pro nor hydroxyproline.

```

CC -1- SUBUNIT: Homodimer.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
 CC a soluble form (By similarity).
 CC -1- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and
 CC several immune system tissues.
 CC -1- PFM: The soluble form (SDP) derives from the membrane form (MDP)
 CC by proteolytic processing (By similarity).
 CC -1- SIMILARITY: Belongs to peptidase family S9B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF461806; AAL67836.1; -
 DR EMBL; AY056834; AAL2628.1; -
 DR MEMPRO; S09_003; -
 DR InterPro; IPR002471; Prol_endopep_ser.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR Pfam; PF00930; DDPV_N_term; 1.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 KM Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;
 KM Transmembrane; Glycoprotein; Signal-anchor;
 FT CHAIN 1 765
 FT CHAIN 38 765
 FT DOMAIN 1 6
 FT TRANSMEM 7 29
 FT DOMAIN 30 765
 FT ACT SITE 629 629
 FT ACT SITE 707 707
 FT ACT SITE 739 739
 FT CARBOHYD 84 84
 FT CARBOHYD 91 91
 FT CARBOHYD 149 149
 FT CARBOHYD 218 218
 FT CARBOHYD 228 228
 FT CARBOHYD 271 271
 FT CARBOHYD 280 280
 FT CARBOHYD 320 320
 FT CARBOHYD 392 392
 FT CARBOHYD 495 495
 FT CARBOHYD 684 684
 SQ SEQUENCE 765 AA; 88369 MW; E32165421F43E116 CRC64;

Query Match 12.0%; Score 566; DB 1; Length 765;
 Best Local Similarity 27.0%; Pred. No. 1,5e-33;
 Matches 191; Conservative 105; Mismatches 274; Indels 138; Gaps 29;

QY 206 PADPMI-----AFIHSNDIWSNVTREERLTIVNHEMLAMERDASAGATVY 256
 DB 148 PNNWTITSSVGHKLAIVWMDIYKXNPNSBSQITWTG-----KDVITNGITDGV 201
 QY 257 IQEE-FDRYSGYWCPKAEETPSGGKILRLIYEENDESEVETIHH--VTSPLMETFRADS 312
 DB 202 YEEEFVSAVSALMWSNSSTF-----LAVAQFNDDVPLPIEVSFYSDSELQPKYTK 252
 QY 313 FRIYKGTANPKVTKMSIMIDABGRITIDVIDKELIOPELLFEGVEYIARAGWTPECK 372
 DB 253 IYPPAGAVNP--TIKFVWNISISPINATSOQVPPGSVLI-GDHYLCVTVWTEER 309
 QY 373 YAMSILDRSQTRLQIVLISPELFIPEBDVMERQRLISVDSVTPLIYEBETDIWIN 432
 DB 310 IS-----LQMLRIQNY--STNDICDYDRSGRWIS 338
 QY 433 IHDHVFVPSHSEET-----EFIFASECKTGFRHLYKITSILKSKYKRSRG 480

DB 339 SVGRCHI-----EISTGWGRFPAEPHPTSDGNS----FYKIS--NEEGYKI-- 383
 QY 481 GLPAPDFKCPKIEEIAITSGEWEVLGRHGSNIQVDEVRLLYFEGT-KDSPLEHLLYV 539
 DB 384 -----CHFQDRKNCFCFITKGAMEVIG-----IEALTSDVLYIISNRYKQMPGRNLKYI 433
 QY 540 SYNPBEVYRLT-----DREYSHSCCISQCHDFEISKYSQKQNPCHVSLYKLSPEDDPT 594
 DB 434 QLNDYKVTCTSELNPNDRCOQYVSFSQEAQYQRLCSGPGLP---LYTLHNSND-- 487
 QY 595 CKTEFWATIDAGPLPD-----YPPPEIFSESTTGFLT-XGMLYKPHDLQPKRYPT 648
 DB 488 --KEL--RYLENNSLDQVLTQVQMPKSLDPIHLLHGYKFWQMTLPPH-PDKSKRYPL 541
 QY 649 VLFYIGAPQVQVQVNNRFKGVKFRNLN---TLASIGYVVV-INDRSGCHGLKFEAGFY 704
 DB 542 LIEVYAGPCSQ-----KDAIFRLWATYLAATENIIVASFGRRSGYGQDKIMHINR 595
 QY 705 XMQQIEIDDOVEGLQYTLASRYPTIDLDVGIHGMSYGGYISLWALMQRSDI FRYAIA GAP 764
 DB 596 RLQTFEVEDQIEATRFQ-SKMGFVDDKRIALMGMSYGGYVTSVVLGAGSGVFKGIAVAP 654
 QY 765 VTLMIYFDGYTERRVWG--HPDNEGGYIIGSVAMQAEKRPSPENLLHGFIDENYVF 822
 DB 655 VSKMEYDYSVYTBKRYGLPPEQNDLSYRSTVMSRAENF--KQVEYLLTHGPAQDVHF 712
 QY 823 AHTSILLFLYRAKREYDQIYQERHSIRVPSGGEHYELHLAYIOE 870
 DB 713 QQAQISKALVDAGVDVDFQSMWVYTDDEHGTLASSTRAHQHITVHMSHFLKQ 760

RESULT 3
 ID DEP4 HUMAN STANDARD; PRT; 766 AA.
 AC E27487;
 DT 01-APR-1992 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
 antigen CD26) (TPI03) (adenosine deaminase complexing protein-2)
 DS (ADABP).
 GN DPP4 OR ADP2 OR CD26.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 RN NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=92329551; PubMed=1352704;
 RA Mishumi Y., Hayashi Y., Arakawa F., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of human dipeptidyl
 RT peptidase IV, a serine proteinase on the cell surface.";
 RL Biochim. Biophys. Acta 1131:333-336 (1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95012454; PubMed=7927537;
 RA Abbott C.A., Baker E., Sutherland G.R., McCaughan G.W.;
 RT "Genomic organization, exact localization, and tissue expression of
 RT the human CD26 (dipeptidyl peptidase IV) gene.";
 RL Immunogenetics 40:331-338 (1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=peripheral blood;
 RX MEDLINE=92325476; PubMed=1352530;
 RA Tanaka T., Camentzi D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,
 RA Dahlberg H.N., Schlossman S.F., Morimoto C.;
 RT "Cloning and functional expression of the T cell activation antigen
 RT CD26.";
 RL J. Immunol. 149:481-486 (1992).
 RN [4]
 RP ERRATUM.

RA MEDLINE=931171637; PubMed=8094732;
RA Tanaka T.;
RL J. Immunol. 150:2090-2090(1993).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=22388257; PubMed=12477931;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA DiCicco L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stappleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Tosiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mulhany S.J.,
RA Boeak S.A., McEwan P.U., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.V., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Witting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Binkley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield J.S.N., Kravitz M.I., Skalska U., Smalins D.E.,
RA Schenck A., Schein J.E., Jones S.J.W., Makra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 1-551 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=92165847; PubMed=1347043;
RA Darnoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,
RA Triotot P., Barbé A.;
RT "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like
RT colon cancer cell lines HT-29 and Caco-2. Cloning of the complete
RT human coding sequence and changes of dipeptidyl peptidase IV mRNA
RT levels during cell differentiation."
RL J. Biol. Chem. 267:4824-4833(1992).
RN [7]
RP SEQUENCE OF 545-766 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=91024044; PubMed=1977364;
RA Darnoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;
RT "Isolation of a cDNA probe for the human intestinal
RT dipeptidyl peptidase IV and assignment of the gene locus DPP4 to
RT chromosome 2."
RL Ann. Hum. Genet. 54:191-197(1990).
RN [8]
RP SEQUENCE OF 1-31 FROM N.A.
RX MEDLINE=96067599; PubMed=7487939;
RA Boehm S.K., Gum J.R., Jr., Erickson R.H., Hicks J.W., Kim Y.S.;
RT "Human dipeptidyl peptidase IV gene promoter: tissue-specific
RT regulation from a TATA-less GC-rich sequence characteristic of a
RT housekeeping gene promoter."
RL Biochem. J. 311:835-843(1995).
RN [9]
RP PARTIAL SEQUENCE.
RC TISSUE=Kidney;
RX MEDLINE=93210468; PubMed=8096237;
RA Morrison A.N., Vijayasaradhi S., Engelstein D., Albino A.P.,
RA Houghton N.N.;
RT "A marker for neoplastic progression of human melanocytes is a cell
RT surface ectopeptide."
RL J. Exp. Med. 177:1135-1143(1993).
RN [10]
RP FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline. Plays a role in T cell activation
CC -1 CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -1 SUBUNIT: Homodimer or heterodimer with seprase (FAP).
CC -1 SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form.

```
CC -|- PMM the soluble form (SDPP) derives from the membrane form (MDPP)  
CC by proteolytic processing.  
CC -|- SIMILARITY: Belongs to peptidase family S9B.  
CC -|- DATABASE: NAME=PROW; NOTE=CD guide CD26 entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd26.htm".  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
DR EMBL; U13735; AAB60646.1; -.  
DR EMBL; U13710; AAB60646.1; JOINED.  
DR EMBL; U13711; AAB60646.1; JOINED.  
DR EMBL; U13712; AAB60646.1; JOINED.  
DR EMBL; U13713; AAB60646.1; JOINED.  
DR EMBL; U13714; AAB60646.1; JOINED.  
DR EMBL; U13715; AAB60646.1; JOINED.  
DR EMBL; U13716; AAB60646.1; JOINED.  
DR EMBL; U13717; AAB60646.1; JOINED.  
DR EMBL; U13718; AAB60646.1; JOINED.  
DR EMBL; U13719; AAB60646.1; JOINED.  
DR EMBL; U13720; AAB60646.1; JOINED.  
DR EMBL; U13721; AAB60646.1; JOINED.  
DR EMBL; U13722; AAB60646.1; JOINED.  
DR EMBL; U13723; AAB60646.1; JOINED.  
DR EMBL; U13724; AAB60646.1; JOINED.  
DR EMBL; U13725; AAB60646.1; JOINED.  
DR EMBL; U13726; AAB60646.1; JOINED.  
DR EMBL; U13727; AAB60646.1; JOINED.  
DR EMBL; U13728; AAB60646.1; JOINED.  
DR EMBL; U13729; AAB60646.1; JOINED.  
DR EMBL; U13730; AAB60646.1; JOINED.  
DR EMBL; U13731; AAB60646.1; JOINED.  
DR EMBL; U13732; AAB60646.1; JOINED.  
DR EMBL; U13733; AAB60646.1; JOINED.  
DR EMBL; U13734; AAB60646.1; JOINED.  
DR EMBL; M74777; AAA51943.1; -.  
DR EMBL; BC013329; AAH13329.1; -.  
DR EMBL; M80536; AAA52308.1; -.  
DR EMBL; X60708; CAA43118.1; -.  
DR EMBL; S79876; AAB35614.1; -.  
DR PIR; S24313; CDHU26.  
DR PDB; INIM; 07-JAN-03.  
DR MEROPS; S09.003; -.  
DR GeneW; HGNC:3009; DPP4.  
DR MIT; I02720; -.  
DR GO; GO:004274; F:dipeptidyl-peptidase IV activity; TAS.  
DR GO; GO:0005209; F:Plasma protein; TAS.  
DR GO; GO:0006955; F:Immune response; TAS.  
DR InterPro; IPRO02469; DPPIV N term.  
DR InterPro; IPRO01375; Peptidase S9.  
DR InterPro; IPRO02471; Prol_endopp_ser.  
DR InterPro; IPRO03379; Ser_estrs_site.  
DR Pfam; PF00930; DPPIV_N_term; 1.  
DR Pfam; PF00326; Peptidase_S9; 1.  
DR PROSITE; PS00708; PRO_ENDOPP_SER; 1.  
KW Hydroxylase; Aminopeptidase; Dipeptidase; Serine processae;  
KW Transmembrane; Glycoprotein; Signal-anchor: 3d-structure.  
FT CHAIN 1 766  
FT FT  
FT CHAIN 39 766  
FT FT  
FT DOMAIN 1 6  
FT TRANSSEM 7 28  
FT FT  
FT DOMAIN 29 766  
FT ACT_SITE 630 630  
FT ACT_SITE 708 708  
FT ACT_SITE 740 740
```

```

FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 685 685 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 6 6 K -> R (IN REF. 6).
FT CONFLICT 7 7 V -> I (IN REF. 1).
FT CONFLICT 437 437 S -> I (IN REF. 1).
FT CONFLICT 557 557 T -> I (IN REF. 7).
FT CONFLICT 663 663 D -> E (IN REF. 7).

```

Query Match 11.2%; Score 525; DB 1; Length 766;
 Best Local Similarity 26.3%; Pred. No. 1.5e-30;
 Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

```

QY 206 PADPDW-----AFHSNDWISNIVYRERRLITYHNEELANNEEDARSGVATFV 256
DB 149 PNTQWTWSPVGHKLAIVYVNDIYKIEPNPSYRTWTG-----KEDIYVGIWDV 202
QY 257 LOEB-FDRYGYWCPKAEETPSGKILRIYENDESEYRIH---VTSPLRETRADS 312
DB 203 YEEVFSAVYALWSPNGTF-----LAVAFNDTEVPLIYSFYSDESLOPKTVR 253
QY 313 FRIPKGTANPKYTFKMSIEMIDAEGRIDVIDKELIQPEILFEGEYIARACWTBEGK 372
DB 254 VPYKAGAVNPVTFVY--VNTDLSSTVNTATSIOIAPASMLI-GHYLCDVYWA---- 306
QY 373 YANSILDRSOTRQIVLISPELFIPEVDVMEQRILIESVPDSVYPLIYEETDWIN 432
DB 307 -----TOERISLOWL-----RRIQNY--SVMDICDYDESSGRW-N 338
QY 433 IHDIHFVFPSSHIEIFIFASECKTGRH-----LYKTSILSKSKYRSQGPAP 485
DB 339 C-----LVARQHIEMSTTGWGRPRSEPHITLDGNSFYKILS--NEGRYRHT----- 384
QY 486 SDRKCPKIEIATISGEWEVLGRHNSIOVDEARLYFEGT-KDSPLEHLLVYVSYNP 544
DB 385 CYFOIDKDKCTFLIKGTWEVYG-----IEALTSPLYIISNEYKMGGRNLVYKI----- 434
QY 545 GEVTRILDRGRVSHSCIS-----QHCDFFISKYNSQCNPH---C-----VSLYKLSPEDD 592
DB 435 -----QUSD-YTKYTCISCEINPERCOYVSFSFEKAKYQLRCSGGLPLYLTHSSVND 488
QY 593 PTCKYEFMATIIDSAGPLPDYTPPE---IFSESTIGFTLYGMLYKPHLOPKKYPTV 649
DB 489 KGLRVLED-NSALDKM--LQNVQMPSKKLPFIILNETKF-WYQMLLPFH-FDSKKTPL 543
QY 650 LFIYGGQVQLVNNRFQVYKFRIN--TLASLGIVVVV-IDNNGSCHRGKTEGAEKFK 705
DB 544 LDVYAGCSQ-----KADTVFRLNWTATYLASTENITIVASFDDGSGCYQSGIKMHAINRR 597
QY 706 MGQIEIDDOVEGLQYLASRYDFTLDVGHGMSYGYLSIMLMQSDIERVALNAPV 765
DB 598 LGTFPEVDDQLEAARQF-SKMGFVDNKRITAIWMSYGYVSWVJGSSGYFKGIAVAP 656
QY 766 TLMIFDYGTETRYMG--HPDQNEQGYLIGSVAMQAEKFPSEBPNRLILHGLDENYFA 823
DB 657 SRREYDYSYTERYMGFPEEDMLDHRNSTVMSRANF--KQVEYLLIHGTADNVHFG 714
QY 824 HTSILISFLVRACKPYDLOIYPORRHSIRVPEGEHEYLHLVLOE 870
DB 715 QSAQISKALVDVGFQAMWYTDDEHGIASSTAHOHLYTHMSHFIKQ 761

```

RESULT 4
 ID DP4_RAT STANDARD; PRT; 767 AA.
 AC P14740;
 DT 01-APR-1990 (Rel. 14, Created)

```

DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26) (SP110 glycoprotein) (Bile canalliculus domain-specific
DE membrane glycoprotein).
GN DP4 OR CD26
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=69122496; Pubmed=2563382;
RA Ogata S., Misumi Y., Ikehara Y.,
RT "Primary structure of rat liver dipeptidyl peptidase IV deduced from
RT its cDNA and identification of the NH2-terminal signal sequence as
RT the membrane-anchoring domain."
RL J. Biol. Chem. 264:3596-3601(1989).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88068516; Pubmed=3479775;
RA Hong W., Doyle D.,
RT "cDNA cloning for a bile canalliculus domain-specific membrane
RT glycoprotein of rat hepatocytes."
RL Proc. Natl. Acad. Sci. U.S.A. 84:7962-7966(1987).
[3]
RN [3]
RP SIGNAL-ANCHOR.
RX MEDLINE=90338089; Pubmed=1974258;
RA Hong W., Doyle D.,
RT "Molecular dissection of the NH2-terminal signal/anchor sequence of
RT rat dipeptidyl peptidase IV."
RL J. Cell Biol. 111:323-328(1990).
[4]
RN [4]
RP SEQUENCE OF 281-302.
RX TISSUE=Kidney;
RX MEDLINE=94128239; Pubmed=7905271;
RA Iwaki-Egawa S., Watanabe Y., Fujimoto Y.,
RT "N-terminal amino acid sequence of the 60-kDa protein of rat kidney
RT dipeptidyl peptidase IV."
RL Biol. Chem. Hoppe-Seyler 374:973-975(1993).
CC -1- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
CC POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
CC PENULTIMATE RESIDUE IS PROLINE.
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
CC A SOLUBLE FORM.
CC -1- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)
CC BY PROTEOLYTIC PROCESSING.
CC -1- SIMILARITY: Belongs to peptidase family 99B.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
DR EMBL; U04591; AAA41096.1; -.
DR EMBL; J02997; AAA41272.1; -.
DR PIR; A39914; A39914.
DR MEROPS; S09.003; -.
DR InterPro; IPR002469; DPPV_N term.
DR InterPro; IPR001375; Peptidase_59.
DR InterPro; IPR002471; Pro1_endopep_ser.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPV_N term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
DR Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;

```

```

KM Transmembrane: Glycoprotein, Signal-anchor.
FT CHAIN 1 767 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT CHAIN 37 767 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT CHAIN 281 767 DIPEPTIDYL PEPTIDASE IV 60 KDA SOLUBLE
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 7 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 29 767 EXTRACELLULAR (POTENTIAL).
FT ACT SITE 631 767 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 709 767 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 741 741 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 686 686 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 38 38 A -> R (IN REF. 2).
FT CONFLICT 183 183 I -> T (IN REF. 2).
FT CONFLICT 332 332 C -> V (IN REF. 2).
FT CONFLICT 352 352 V -> D (IN REF. 2).
FT CONFLICT 394 394 L -> F (IN REF. 2).
FT CONFLICT 562 562 R -> Q (IN REF. 2).
FT CONFLICT 624 624 R -> Q (IN REF. 2).
SQ SEQUENCE 767 AA: 88003 MW: AA17C0EC6F0E4652 CRC64:

Query Match 11.1%; Score 523.5; DB 1; Length 767;
Best Local Similarity 26.0%; Pred. No. 28-30; Indels 165; Gaps 33;
Matches 188; Conservative 104; Mismatches 266;

QY 206 PADPWI-----AFHSNDIWSNIVTREERLTIVHNEILNMEEDARSAGVATFV 256
DB 147 PNNQWIMSGEGHKLAYVNDIYKIBPHPSHRIT-----STGENVIFNFINQWV 200
QY 257 LOEF-FDRISGYWCPKAEITPGSGKIIRILEENDESEVEIIHVTSPMLETRADSRF 315
DB 201 YEEELFGAYSLMW-----SPNGTFLLAYAFNDGVPLIEISFYSDLSLQPKTWIPIY 254
QY 316 PRTGAPKATPFKMEIMIDAGRILIDVIELIOPFELFEGVEYIARAGTPEPKXAW 375
DB 255 PKAGVAVPTVFFI--VNTDLSSTTTTIPMOITPASYT-TGDHYLCDAVAVSDRIS- 310
QY 376 SILDRSQTRLQIVLISPELFIPEVDYMERQRLIESYVDSVTPLIYEETDIWINTHD 435
DB 311 -----LQWLRRIQNY--SYMAICDVKITLVW-NC-- 337
QY 436 IFHVPQSHERIEFIIPASFC-----KTGRRLKYITSLIKE 472
DB 338 -----PTQSHIEISATGCGRRPAPHPHTSDSSFYKTVSDQGYKH--ICQFOKD 388
QY 473 SKYRSSGGPLPAPSPFKCPKEEIA--ITSGEWEVLGRHSGNIQVDERRLVYEPGT-KD 529
DB 389 RK-----PEQVTFITTKAMEIT-----SIEMLTSDIYLTISNEYKE 425
QY 530 SPLEHLLIVSYNVEGAVTLTRDGYSHSCIS-----QHCDFF--ISKTSNOKNPHC- 580
DB 426 MFGGNLYKI-----QLTD--HTMKKCHSCLNBERCQYVSLSKEAKYQLQCR 474
QY 581 ---SVLYLSSPEDPTCKTEFMAITILDSAGPLPDYPPR---IFSESTTGFLYLM 634
DB 475 GPGLEPLYLHRSTQCKELRVED-NSALDKM--LQDVOMPSKDLFIYLNTRF-WYOMI 530
QY 635 YKPHDLQPGKKYPTVLFTYGGPOVQVANNRFGKGYVFFLN--TLASGYVVV-INDRG 690
DB 531 LPPH-FDKSKYPLIIDVYAGCSG-----KADAPFLNMTATYIASFENIIVASFDRG 583
QY 691 SCHGGLKREGAFKTKMGQIEIDQVEGL-QYIASYDITIDRGIHGMSTGGYISLMAL 749

```

```

DB 584 SGYQGDKIMAHINKRLGLEVEDQIEAARQFL--RMGEVDSKRAVIKMGSTGGYTSWL 641
QY 750 MQRSDIFFEVALAGAPVTLMIFFDYGTGYRNG--HPDQNEQGYVLSGYAMQAEKPPSPN 807
DB 642 GSGSVFECGIAVAPVSMWEYDYDSYTERVWGLPPEPNDLHNRSTMSARENF--KQV 699
QY 808 RLLIHGFLDENVHPARTSILSLFVRAGKPYDILQIYQERHSIRVPSGHEYLHLIHY 867
DB 700 EYLLIHGTADNVHVFQQAQISKALVDAGVDFQAMWYDEBDHGIASTAHQIYSHMSHF 759
QY 868 LOE 870
DB 760 LQQ 762

RESULT 5
YEAS SCHPO STANDARD: PRT; 793 AA.
AC 014073;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative dipeptidyl aminopeptidase C2E11.08 (EC 3.4.14.-).
GN SPACUN4.08 OR SPAC2E11.08 (Fission yeast).
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=2184601; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voiclaert G., Aert R., Robben J., Grymoprez B.,
RA Welljens I., Vanstraelen E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Dada R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shipukovskii G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe." ;
RL Nature 415:871-880(2002).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
CC vacuoles (By similarity)
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL031180; CAA20138.1; -
CC PIR; T41703; T41703.

```


DR Genedb_Spombe; SPACUNK4.08; -
 DR InterPro; IPR002469; DDPiv_N term.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR002471; Pro1_endopep_ser.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR Pfam; PF003930; DDPiv_N term; 1.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 KW Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
 KW Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
 FT DOMAIN 1 24
 FT TRANSMEM 25 45
 FT (POTENTIAL-ANCHOR (TYPE-II MEMBRANE PROTEIN))
 FT LUMENAL (POTENTIAL).
 FT ACT_SITE 647 793
 FT ACT_SITE 722 722
 FT ACT_SITE 755 755
 FT CARBOHYD 101 101
 FT CARBOHYD 136 136
 FT CARBOHYD 246 246
 FT CARBOHYD 299 299
 FT CARBOHYD 303 303
 FT CARBOHYD 324 324
 FT CARBOHYD 336 336
 FT CARBOHYD 377 377
 FT CARBOHYD 384 384
 FT CARBOHYD 407 407
 FT CARBOHYD 535 535
 FT CARBOHYD 761 761
 FT SEQUENCE 793 AA; 91304 MW; 208707F97F231463 CRC64;

Query Match 11.1%; Score 519.5; DB 1; Length 793;
 Best Local Similarity 26.3%; Pred. No. 4e-30;
 Matches 187; Conservative 113; Mismatches 256; Indels 155; Gaps 32;

QY 212 IAFHSNDIMI-----SNIVREBRRLTY-----VHNELMNERDASAGAFVLOEE-F 261
 DB 183 LSFVNNNDIVKKNQNV-----QRLTYGYDVFN-----GLTWIVYEEVL 225
 QY 262 DRYSGYWCPCAKETTPSGGKILRIIYEENDESEVELIHV-----TSPMLSTRADS 312
 DB 226 SSPSTIWMSPDSD-----KIAPLKNESEIPYHYPLVYALDELPLPEDVYNKMA 276
 QY 313 FRYPKGTANPKVTFKMSIMIDAEKRIIDVIDKELIOPELLFBEVYVIAAGWTEBCK 372
 DB 277 IYPPKGNPNPVSLEFADLNSNASSNFMENPELAEP-----VVOVVLVWVNS----- 326
 QY 373 YAMSILL-----DRSOTFLQIVLISPEL--FIPVEDVMER-----ORLISVPDSVTPLLI 422
 DB 327 ---SVIVQFTNNNSTCITARLDELTKSIHYKTECLEGWEVQASAKMP--LNSILV 381
 QY 423 YEETDIWNIHIDIFHVPQSHBEIEFLFASECKTGFRHLYKITSILKSKYKSSGGL 482
 DB 382 WENMSD--GYFDILALDDYNHLAFIPF-----NGSS 410
 QY 483 PAPSPKCIKEKIEIATSGEWEVLGHSNIOVDYERKLYVEGTYDSELEHLLVYVYV 542
 DB 411 P-----IYVTSAMVLT--DGRPHIDGFGNVYPLATLKOSTERHLYVVS-L 454
 QY 543 NGEVTRILDRGYS--HSCCISOHCDPEISKYSQKNCHCVSLYKLSPEDEPCKTKE 599
 DB 455 DLEIYGITDNEDSGYTSFSPRODFVLVNHGDVY---WQELRKTOKDYCLSLIE 510
 QY 600 FWATILDSAGPLPDYTPPEI---FSFESTGFTLYGMYLKEPHDLOPKKYPVTLFTYGG 655
 DB 511 TNSRLKQO---LSSITLPSVEYKLTENDTL--FNF--MERRRPNPDVNNKYPVLFFAYGG 564
 QY 656 PVOVLVNNRFKGVKFERLNTLASLG-----YVVVVINDRGSGCHGILKKEGAFYKMGQI 709
 DB 565 PEGQGV-----AKIFRVDFOAYILASHDFEFTVTLDERGFGNGNARFYSVSSHLLGW 618
 QY 710 EIDOVEGLQYLAGRYDFIDLRVGIHGSYGYSLMALMQRSDIFRVAIAGAPVTLMI 769

DB 619 ESYDQQAQAKFMAD-LPFPDENHVGIGWMSYGYTLTKLT-ETQDVFSYGMAVAPVTDNR 676
 QY 770 FYDTGYTERYMGHPDNEQGYLYGVAMQAEKFPSEPNLLHAGFLDENVHFAHTSILL 829
 DB 677 LVDSVYTERYMDLPQYNKSG-YKNSQIDHYEKF-XQLKXFFVAHGGDDNVHGHSHMLM 734
 QY 830 SFIVRAG-KPYDIQYPERHSIRPESGEHLEHLLHYQENLGSRIAL 879
 DB 735 DGINLANCYNDMAVFPDSAHSI---SYHNASLSIYHRLSEWIGDALGRI 781
 RESULT 6
 ID DEPA_MOUSE STANDARD; PRT; 760 AA.
 AC P28843;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
 DE antigen CD26) (Thymocyte-activating molecule) (THAM).
 GN DPP4 OR CD26.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=Swiss; TISSUE=Thymus;
 RX MEDLINE=92129288; PubMed=1370813;
 RA Marguet D.A., Bernard A.-M., Vivier I., Darmoul D., Naquet P.,
 RA Pierres M.;
 RT "cDNA cloning for mouse thymocyte-activating molecule. A
 RT multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a
 RT subgroup of serine proteases.";
 RT J. Biol. Chem. 267:2200-2208 (1992).
 RN [2]
 RC REVISIONS.
 RA Marguet D.A.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RC SEQUENCE FROM N.A.
 RC STRAIN=B10.A; TISSUE=Liver;
 RX MEDLINE=95092780; PubMed=7999781;
 RA Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;
 RT "Structure of the mouse dipeptidyl peptidase IV (CD26) gene.";
 RL Biochemistry 33:15204-15214 (1994).
 RN [4]
 RC SEQUENCE FROM N.A.
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haich P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin A.G., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M., Umed T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mollahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs B.L.,
 RA Rahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [5]
 RC SEQUENCE OF 1-20.
 RX MEDLINE=91302787; PubMed=1712807;
 RA Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,

RA Bernard A.-M., Gouvel J.-P., Pierres M.;
 RT "evidence that thymocyte-activating molecule is mouse CD26
 (dipeptidyl peptidase IV).";
 RL J. Immunol. 147:447-454(1991).
 CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
 CC polypeptides having unsubstituted N-termini provided that the
 CC penultimate residue is proline.
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-
 CC xcc, from a polypeptide, preferentially when Xbb is Pro, provided
 CC xcc is neither Pro nor hydroxyproline.
 CC -1- SUBUNIT: Homodimer.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
 CC a soluble form.
 CC -1- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
 CC by proteolytic processing.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X58384; CAA41274.1; -.
 DR EMBL; U12620; AAA82213.1; -.
 DR EMBL; U12599; AAA82213.1; JOINED.
 DR EMBL; U12600; AAA82213.1; JOINED.
 DR EMBL; U12601; AAA82213.1; JOINED.
 DR EMBL; U12602; AAA82213.1; JOINED.
 DR EMBL; U12603; AAA82213.1; JOINED.
 DR EMBL; U12604; AAA82213.1; JOINED.
 DR EMBL; U12605; AAA82213.1; JOINED.
 DR EMBL; U12606; AAA82213.1; JOINED.
 DR EMBL; U12607; AAA82213.1; JOINED.
 DR EMBL; U12608; AAA82213.1; JOINED.
 DR EMBL; U12609; AAA82213.1; JOINED.
 DR EMBL; U12610; AAA82213.1; JOINED.
 DR EMBL; U12611; AAA82213.1; JOINED.
 DR EMBL; U12612; AAA82213.1; JOINED.
 DR EMBL; U12613; AAA82213.1; JOINED.
 DR EMBL; U12614; AAA82213.1; JOINED.
 DR EMBL; U12615; AAA82213.1; JOINED.
 DR EMBL; U12616; AAA82213.1; JOINED.
 DR EMBL; U12617; AAA82213.1; JOINED.
 DR EMBL; U12618; AAA82213.1; JOINED.
 DR EMBL; U12619; AAA82213.1; JOINED.
 DR EMBL; BC02183; AAH22183.1; -.
 DR MEROPS; S09.003; -.
 DR MGD; MGI:94919; DDP4.
 DR InterPro; IPR002469; DDPV_N term.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR002471; Prol_endopep_ser.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR Pfam; PF00930; DDPV_N term; 1.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 KM Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;
 KM Transmembrane; Glycoprotein; Signal-anchor.
 FT CHAIN 1 760
 FT (MDPP)
 FT DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
 FT CHAIN 37 760
 FT (SDPP)
 FT DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
 FT DOMAIN 1 6
 FT CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 7 28
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 29 760
 FT CYS-RICH.
 FT ACT SITE 295 466
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 624 624
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 702 702
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 734 734
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT 83

FT CARBOHYD 90 90 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 213 213 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 514 514 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 679 679 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 760 AA; 87436 MW; A5F644B6A43DF8 CRC64;
 Query Match 11.0%; Score 517.5; DB 1; Length 760;
 Best Local Similarity 25.0%; Pred. No. 5,3e-50;
 Matches 203; Conservative 107; Mismatches 306; Indels 197; Gaps 37;
 QY 131 MYSEBELLRERRIGTVGI-----ASYDYQSGSTFF-----QAGS 168
 DB 67 LYQENNNILLNAHNSSTFLENSTFESGCHSVSDRLFVLENNYKQMHSTASY 126
 QY 169 GIYHVXKDGPGQFTQGPLRPENLVETSCPNIRMDPKLCPADPMI-----AFHSND 219
 DB 127 NIYDVNK-----KOLITEEKIPN-----NTQWITWSPGHLKLAIVWKND 165
 QY 220 IWSNIVTEERLTYVNHLEAMEEDARSAGVATVLOE-PDRISGYWCKR----- 272
 DB 166 IYKVEPHLPSSRIT-----STGEENVYNGITDWVEEVRGAYSALWWSPPNNTFLAY 219
 QY 273 AETTPSGKILRIIYENDESEVEIITHVTSPLMETRADSFRYPKGTGTANPKYFKMSEI 332
 DB 220 AQFNDTGVPLEYSF-YSDS-----LQYKWTWIPRPKAGAVNPTKFF--V 265
 QY 333 MIDASRIIDVIDKELIOPPEILFEGEYIARAGWTPGKYA-----WSLLDRSQ 383
 DB 266 NIDSLSSSSAPIQIPAPASVA-RGDHYLCVWMAETEIRISLQMLRIQNYGVMAICDY 324
 QY 384 TRQIVLISPELPIYEDDWMERORLIESVPDSVTLLIYEETDWINIHDFHVPQS 443
 DB 325 DKTNLTWNCPS-----EQOHV-----EMSTGTVG--RFRPAEPHF 358
 QY 444 HEEIEIFPASECKTFRLHYKITSILKESKYRSSGGLPAPADF-KCPIKEIATITSGE 502
 DB 359 TSDGSSFFKISIKDQYKHCNR-----PDKKDCITF-----ITKGA 395
 QY 503 WEYVGRHGSNIQVDEVRLVYFEGT-KDSPLEHLLVVSIVNDEVTRLTRDGYSHSCCI 561
 DB 396 WEVY-----SIEALTSDDYLYISNQYKEMEGGNLYKI-----QLTD--HTNVKCL 439
 QY 562 S-----QHCDFFISKXSNQKNPH--C-----VSLYKLSPEDDPTCKTEFMAITIDSAG 609
 DB 440 SCDLNPBRCQYVAVSFSKEAKYQOLCQWGBPLTYLTHRSTDKELRVLE-----DNSA 493
 QY 610 ---PLPDYTPPE---IFSESTGFTLLYGNLYKPHDLQPKKYPYLVFIYGGHVOQLVNN 663
 DB 494 LDBMLQDVQMPKSKLDFIVLNETRF-WYQMLLPFH-FDSKSKYPLLDVYVAGCSQ---- 547
 QY 664 REFQVAKYFRLN--TLASLGIVVVV-IDNRGSHRGILKFGAGAKYKMGQJIEIDQVGLQ 719
 DB 548 --KADASFRLNMAFYASTENIIVASFDDGSGSYQDCKIMHAINRRLGTELEVDDQIEAAR 605
 QY 720 YLASRDFDLDLVGVIHGSVYGYLSLMLIMQSDIFRYAIAAPATLMIFYTGYTERY 779
 DB 606 QPFIKGFVDSKQVAILWGSVYGYTSMLVSGSSGVFKCIAVAPVSRWYIYDSVTERY 664
 QY 780 MG--HPDQEGQYGLASVAMQAEKFPSEPNRLILLHGFIDENVHFAHTSILLSFLYRAG 837
 DB 665 MGLPIEDNLDHYRNSVTWMSRAHF--KQVEYLLIHGTADNVHFGQSAQISKALVDAGV 722
 QY 838 PYLOQIYPOGRHSIRPESGEHTEHLIAYLOE 870
 DB 722 DFQAMWYTDDEHGIASSTAHOHTYSHMSHPLOQ 755
 RESULT 7
 DAP2_YEAST

ID	DAP2_YEAST	STANDARD;	PRT; 818 AA.
AC	P18962;		
DT	01-NOV-1990 (Rel. 16, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DBP B) (YSCV).		
GN	DAP2_OR YHR026C.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
CC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
RN	NCBI_TaxID=4932;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=89174971; PubMed=2647766;		
RA	Roberts C.J., Pohlig G., Rothman J.H., Stevens T.H.;		
RT	"Structure, biosynthesis, and localization of dipeptidyl		
RT	aminopeptidase B, an integral membrane glycoprotein of the yeast		
RT	vacuole."		
RL	J. Cell Biol. 108:1363-1373(1989).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=G288c / AB972;		
RX	MEDLINE=94378003; PubMed=8091229;		
RA	Johnston M., Andrews S., Brinkhan R., Cooper J., Ding H., Dover J.,		
RA	Du Z., Favallo A., Fulton L., Gatung S., Geisel C., Kirsten J.,		
RA	Kucelale P., Louis E.J., Macri C., Johnston L., Langston Y.,		
RA	Nham M., Rifkin L., Riley L., St Peter H., Trevasik E., Vaughan K.,		
RA	Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,		
RA	Vaudin M.		
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome		
RT	VIII."		
RL	Science 265:2077-2082(1994).		
CC	-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSSOME-LIKE		
CC	VACOLES.		
CC	-I- SIMILARITY: Belongs to peptidase family S9B.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as their content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isd-sib.ch/announce/		
CC	or send an email to licenseds@sib.ch).		
CC			
DR	EMBL; X15484; CAA33512.1; .		
DR	EMBL; U10399; AAB68879.1; .		
DR	PIR; SA6780; A30107.		
DR	MEROPS; S09_006; .		
DR	SGD; S0001070; DAP2.		
DR	GO; GO:0000329; C:vacuolar membrane (sensu Fungi); IDA.		
DR	InterPro; IPR002469; DPPIV_N term.		
DR	InterPro; IPR001375; Peptidase_S9.		
DR	InterPro; IPR002471; prol_endopep_ser.		
DR	InterPro; IPR003379; Ser_sstrs_site.		
DR	Pfam; PF00930; DPPIV_N_term; 1.		
DR	Pfam; PF00326; Peptidase_S9; 1.		
DR	PROSITE; PS00708; PRO_ENDOPEP_SER; 1.		
KM	Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;		
KM	Transmembrane; Glycoprotein; Signal anchor.		
FT	DOMAIN	1..29	CYTOLASMIC (POTENTIAL).
FT	TRANSMEM	30..45	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT			(POTENTIAL).
FT	DOMAIN	46..818	LUMENAL (POTENTIAL).
FT	ACT_SITE	679..679	CHARGE_RELAY_SYSTEM (BY SIMILARITY).
FT	ACT_SITE	756..756	CHARGE_RELAY_SYSTEM (BY SIMILARITY).
FT	ACT_SITE	789..789	CHARGE_RELAY_SYSTEM (BY SIMILARITY).
FT	CARBOND	63..63	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOND	79..79	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOND	110..110	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOND	139..139	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOND	372..372	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOND	392..392	N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match	10.3%	Score 484.5	DB 1	Length 818
Best Local Similarity	26.0%	Pred. No. 1,6e-27		
Matches 183	Conservative 95	Mismatches 250	Indels 177	Gaps 35
FT CARBOHYD	421	421	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CARBOHYD	738	738	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CARBOHYD	83	83	Q -> H (IN REF. 1).	
FT CONFLICT	125	125	S -> N (IN REF. 1).	
FT CONFLICT	182	188	PEPTONE -> LRELET (IN REF. 1).	
FT CONFLICT	200	200	D -> N (IN REF. 1).	
FT CONFLICT	356	375	TSNVRANSS -> DFKGKERKF (IN REF. 1).	
FT CONFLICT	808	818	AKRADGQFVK -> QSVLSMGNLTNELLTYSSHRDIHKT	
FT			PSYHTWYI (IN REF. 1).	
SEQUENCE	818 AA	93404 MW	318F450445375BD3 CRC64:	
Query Match	10.3%	Score 484.5	DB 1	Length 818
Best Local Similarity	26.0%	Pred. No. 1,6e-27		
Matches 183	Conservative 95	Mismatches 250	Indels 177	Gaps 35
QY 212 IAFHSNDIMSNITREERLTYHNE---	LANNEEDARAGVATFVLQEE-FDRYSG	266		
DB 201 IAYVDNNNIYISATSKITRA--VTNQGSSFLFNKQPD-----	WYEEVEVEDDKA	250		
QY 267 YWMCCKAETTPGSKILRIIVEENESSEY-ELIHTSPLETRRAD-----	SPFRPKRG	319		
DB 251 AMNSTGQ-----	YLAFLKIDSEVEGET--IPYVODEKDIYPMKSIKYPASG	298		
QY 330 TANPKVTEKMEIMIDAGRIIDVIDKELIQP--FEILFEGVEYIARQWTEPKGYAMSI		377		
DB 299 TPNP-----	HAELIWMYMKDQTSFPHRISGNKKQDGLLTETWTWANG--	343		
QY 378 IL---DRSQTLQIYLISPELFIPEDVDYMERQRLIESVPSVTPLLIYEETDIWINI-	433			
DB 344 LVKTTDRSSDILTYVLI-----	DTIAKTSNVVRN-----	ESSNGGWMEIT	383	
QY 434 HDIFAV-----	PROGHEEIEFIPASECKTGPRHLKYKLTSLKESKYRSGCLPAPBDF	488		
DB 384 HNTLFIPANEFFDRPHNGYDILPIG-----	GNH-----	AYENSNS--	SHY	425
QY 489 KCPIKEELIATISGEMVELGRHGSNIQVDEVARLYVEEGTKDSPLEHHTLVYSVYNPGEYT	548			
DB 426 K-----	TLTEGKWEVY--NGPLAFDSMENRLYFISTRKSTTRHYIYIDLRSPNEII	475		
QY 549 RLTRDGRYS--HSCCLSQHCDPFISKYSNOKNPH-----	C-----	VSLYKL	586	
DB 476 EVTLTSEDGVADVSSFGRRFGLLTYKGPKYOKIYDFHSRKAKEKCDKGNVLGKSLYHL	535			
QY 587 SSPEDDPCKTKEFWATIDAGPLPDYTPPELISFESTGTLYG--MYKPHLOP--	642			
DB 536 EKNE-----	VLTKI--LEDIAVPRKSFRELINLKDEFGKDIIVNSYEILPND	580		
QY 643 -----	GKKYPTVLEFIYGGPOVQLVNNRPKGYKVEFLN---TLASLGVVVVINDRSGCH	693		
DB 581 FDETLSDHYLPFFAFYAGPSNGQV-----	YKTSVGENEVAAGLAIIVVVDKRGKGF	634		
QY 694 RGLKPEGAFFKYMGOEIIDQVEGLQYLASRYFDIDLRVGIHGSYGYLSMALMORS	753			
DB 635 KGQDFRSILVRDLRDLYEARQDISAASLYGS--LTFVDPQKISLFGMSYGGYLLTKLTLEKD	693			
QY 754 D-IRRVALAGPYLWLFYDGYTERVYNGHNDQNGGYLLGSV--	AMQAEKPPSEPNR	808		
DB 694 GRHRYKGSVAPVDWIRFYDSVTERYEMHTPEQENFDGVSSEYHNVATL-----	AQANR	747		
QY 809 LLLHGFLENVAFHAFHSILLSFYVRAG--KPYDQIYPOEHSIR	852			
DB 748 FLHMHGTGDDNVHFQNSLKFLDLDLNGVENVYDVHVPFSDHSIR	792			
RESULT 8				
SEPR_MOUSE	STANDARD	PRT	761 AA	
ID SEPR_MOUSE	SEPR_MOUSE	STANDARD	PRT	761 AA
PT 28-FEB-2003	(Rel. 41, Created)			
DT 28-FEB-2003	(Rel. 41, Last sequence update)			
DT 15-SEP-2003	(Rel. 42, Last annotation update)			
DE Septrae (BC 3.4.21.-)	(Fibroblast activation protein alpha) (Integral			
DE membrane serine protease).				

OS FAP.
 ON Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC STRAIN=BALE/c; TISSUE=embryo;
 RX MEDLINE=97284459; PubMed=9139873;
 RA Niedermeyer J., Scanlan M.J., Garin-Chesa P., Dalber C., Fiebig H.H.,
 RA Old L.J., Rettig W.J., Schnapp A.
 RT "Mouse fibroblast activation protein: molecular cloning, alternative
 RT splicing and expression in the reactive stroma of epithelial
 RT cancer".
 RL Int. J. Cancer 71:383-389 (1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Breast;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M., Soares M.B., Bonaldo M.F., Casavert T.L., Scheetz T.E.,
 RA Stempleton M.J., Usdin T.B., Toshiyuki S., Caminci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
 RA Bask S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultyk S.W.,
 RA Villard D.K., Wuzny D.M., Sodergren E.U., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting R.W., Madan A., Young A.C., Rubin G.M., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Glimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.T., Skalska U., Smalins D.E.,
 RA Schermer A., Schein J.E., Jones S.J.W., Marra M.A.).
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences".
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -1- FUNCTION: May have a role in tissue remodeling during development
 CC and wound healing, and contribute to invasiveness in malignant
 CC cancers.
 CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
 CC type IV collagen, but not native type I or type IV collagen. Does
 CC not cleave laminin, fibronectin, fibrin or casein.
 CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
 CC inactive (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=P97321-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P97321-2; Sequence=VSP_005368;
 CC Name=3;
 CC IsoId=P97321-3; Sequence=VSP_005369;
 CC -1- TISSUE SPECIFICITY: Detected in fibroblasts, in placenta, uterus,
 CC embryos from day 7-19 and in new-born mice (P1).
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Y10007; CAA71116.1; --
 DR EMBL: BC019190; AAH19190.1; --
 DR MEROFS; S09.007; --
 DR MGD; MG1:109608; Fap.
 DR InterPro; IPR002469; DPP4V N term.

DR		InterPro; IPR001375; Peptidase_S9.
DR	InterPro; IPR002471; Prol_endopep_ser.	
DR	InterPro; IPR003379; Ser_estrs_site.	
DR	Pfam; PF00930; dPEPiv_N term; 1.	
DR	Pfam; PF00326; Peptidase_S9; 1.	
DR	PROSITE; PS00708; PRO_ENDOPEP_SER; 1.	
KM	Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor; Glycoprotein; Alternative splicing.	
KM	Glycoprotein; Alternative splicing.	
FT	DOMAIN	1 4
FT	TRANSMEM	5 25
FT		
FT	DOMAIN	26 761
FT	ACT_SITE	624 624
FT	ACT_SITE	702 702
FT	ACT_SITE	734 734
FT	CARBOHYD	49 49
FT	CARBOHYD	92 92
FT	CARBOHYD	99 99
FT	CARBOHYD	227 227
FT	CARBOHYD	314 314
FT	CARBOHYD	679 679
FT	VANSPPLIC	31 35
FT		
FT	VANSPPLIC	31 63
FT		
FT	CONFLICT	737 737
FT	SEQUENCE	761 AA; 87944 MW; 9174C3AEDBA21B25 CRC64;
SQ		
 Query Match 10.2%; Score 478.5; DB 1, Length 761; Best local similarity 22.3%; Pred. No. 3,9e-27; Matches 181; Conservative 133; Mismatches 293; Indels 205; Gaps 311		
OY	129 YGWSREBELLRERRIGTVGIASY-----DYHGSGRGFLFQAGSGIYVKD	175
Db	79 YNIERESYIIISNSTKMSVNAITDYGLEPDROPVYLESDYSK-LMYRYTATYYIDLQN	137
OY	176 GSGPGFTQPLRPNLIVETSCPNIRMDPKLCPADPM-----IAFIHNDIWISNTVRE	229
Db	138 G---EFVAGVELPRPIYLIC-----WSPVGSKLAYVYQNNIYIKORPGDP	179
OY	230 ERLITYV--NHELAMMEDASAGATVLOEFL--DRISGYWMCKAETTSGGKILRI	285
Db	180 PFQITLYGREKRIFN-----GIPDWYEEEMLATKYA-LMWSP-----DGKFL--	221
OY	286 LYENDESEVEIHHT-----SPLTEFRADSFYPKGTANPKYTFKSEIMIDAEGR	339
Db	222 AYVEFNDSDFIIAVSYGDQGP-----RTINIIYPRGAGANGPYRV-----F	265
OY	340 IIDVIDKLIOPELLFGVEVIYAAGCTPBCKIAMSILLDSQRLLQIVLISPBLFIV	399
Db	266 IVDTTYPHNVMEVPV--EMIASDY----YFMULTWVSSEVRCLQWL-----	309
OY	400 EDVWERORLIYESVDSTPLIITYETDWINIHIDIFY-----PQSHE	445
Db	310 -----KRWQNV--SVLSICDPREDHANECPRNQEHVESRKGMAGCFVSTPAESQ	359
OY	446 ELIPIFASBECKTGFRHLKYITSLIKESKYRSSGGLAPSDFKCPKEKEIATSGEWV	505
Db	360 DATSYKKIIFSDDKGXKHIIHY-----KDTVENAIQTTSCKMEA	397
OY	506 LGRHSNIQVDEVARLVYEESTKDS-----PLEHHLYVVSYNVPGEVTRLTDG	554
Db	398 -----IYIFRVTLQDSLFSNSNEBGYPGRNRNYRISIGNSPSRK-----	437
OY	555 YSHSCISQH-----CDFIFISKYNSOKNPCHVSILY-----KLSSPEDDPCKCKEFAWIL	605
Db	438 -----CVTKHRKRCQQYTASFYKAYVALVCYGFGLPISTLHDGRDQGIQVLEBK	492
OY	606 DSAEPDLDPYPPPI-FSESFTTGCTLYMLYKHPDLOGCKKYPTVLFIYGGQVQLVANK	664
Db	493 ELENSLRNIQIQPKVEIKKLKDGGGLTFWKMLLPQPDFRSKRYPLLIIQVYGGCSOSVKS-	551
OY	665 FKGVKVFRLNTIALI-----GYVVVVVINDRSGCHGLKEGAFKYMGQIEIIDOVGELOY	720

```

Db 552 -----VAAMWITTLAKKEGIVIALVDORGAFCGDKFLHAHYKRLGYEVEDQTLAVRK 606
Qy 721 LASYDEIDLDRAVGHMSYGYSLMALMQRSDIFRAVAGAPVTLIMFYDTGYTERYM 780
Db 607 FL-EMGFIDEERLAIWMSYGYVSSIALAGTGLPKCGIAPVAVSSMEVYASISERFM 665
Qy 781 GHP--DQNEQYIYGSVAMQAEKPPSEPNRLLLHGFIDENAFHATSILISFLVRAKPP 838
Db 666 GLPKDNLLEHYKNSTVWARAEYFRND--YLLIHGTADNVHFNQNSQIALVNAQVD 723
Qy 839 YDLQYPERHSIRVPESGEHYELHLHYLOE 870
Db 724 FQAMWYSDQNHGISGSSQNHLYTHMTHFLKQ 755

RESULT 9
ST13_YEAST
ID ST13_YEAST STANDARD; PRT; 931 AA.
AC P33894;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase A (EC 3.4.14.-) (DPAp A) (YSCIV).
GN STE13 OR YCII OR YOR219C OR YOR50-9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9506382; Pubmed=7975897;
RA Anna-Artiola S.S., Herskowitz I.;
RT "Isolation and DNA sequence of the STE13 gene encoding dipeptidyl
RT aminopeptidase."
RL Yeast 10:801-810 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=AB320;
RA Flanagan C.A., Thorner J.;
RT "STE13."
RL (in) Getting M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.);
RL Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
RL Press, Oxford (1994).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=5288C / FY1679;
RX MEDLINE=9643797; Pubmed=8840505;
RA Galisson F., Dujon B.;
RT "Sequence and analysis of a 33 kb fragment from the right arm of
RT chromosome XV of the yeast Saccharomyces cerevisiae."
RL Yeast 12:877-885 (1996).
CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
CC ALPHA-FACTOR PRECURSOR.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSGOSOME-LIKE
CC VACUOLES.
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L21944; AAA35119.1; -
DR EMBL; U08230; AAA17897.1; -
DR EMBL; X92441; CA53182.1; -
DR EMBL; Y75127; CAA99437.1; -
DR PIR; A49737; A49737.
DR MEROPS; S09.005; -
DR SGD; S0005745; STE13.

```

```

DR GO; GO:0005802; C:Golgi trans-face; IDA.
DR GO; GO:0004177; F:aminopeptidase activity; IDA.
DR GO; GO:0007323; P:peptide hormone maturation; IDA.
DR InterPro; IPR002469; DPIP N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Prol_endopep_ser.
DR Pfam; PF00930; DPIP N term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; FALSE NEG.
KM Hydrolyase; Aminopeptidase; Dipeptidase; Serine protease;
KM Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.
FT DOMAIN 1 119
FT TRANSMEM 120 140
FT DOMAIN 141 931
FT ACT_SITE 785 785
FT ACT_SITE 863 863
FT ACT_SITE 896 896
FT CARBOHYD 377 377
FT CARBOHYD 814 814
SQ SEQUENCE 931 AA; 107200 MW; 81AF70094093C023 CRC64;

```

```

Query Match 9.9%; Score 465; DB 1; Length 931;
Best Local Similarity 21.5%; Pred. No. 5.1e-26;
Matches 218; Conservative 150; Mismatches 38; Indels 268; Gaps 41;

```

```

Qy 6 ETEQLGEIPEFETADCEENIESODRPKLEPVERYSQJLKLADTRKHYGMMAKAPH 65
Db 43 QSNENGTINIRPTEITIVTDVPC-TPLQEQYS-----WRRRE 83
Qy 66 DFMFVKANDDPGPHSDRIYLAWS-----GENRENTLFYSEIPKTIINRAVILMS 115
Db 84 SPQE---NDIENQHTHSFVSNAKFNRWGEMSLPEKRSVIVVTLALASVVLVLILP 140
Qy 116 WKPL-----LDLFOATL-----DYGM 131
Db 141 SKLLPKTKRPTKSAGDSSLGKRSFSLFNVLNGDFALPEDTFHFIDPQRLGQDSBGL 200
Qy 132 YSRBEELRERKRIG-TVGIASDYHQSGTFLRQ----- 165
Db 201 YFTTKELDHTNFIACLPETFEVUNGKRFYEGVEFVSIVQVINYKLDKILFTNLE 260
Qy 166 ----AGSGIYHVKDGPQGFQOPLRPNLVETSCPNIRMDPK-----LCPAD 208
Db 261 SEFRHSSKGFYWKD-----INTGNIERPLPEKSDDNVELGLSKLSVAHSPRA- 309
Qy 209 EDWIAFIHSDIATISNIVTREERLTY-VNENLANNEEDARSAVATFVIOEE-FDRYSG 266
Db 310 YNIVYFVENNLFLQOVNSGVAKVTEDGSKDIFNAKPD-----WVEEVLASDQA 361
Qy 267 YWMCPRAEETPSGKILRLIYENDESEVEIIVTSPMLSTRADS-FRYPKGTANPKV 325
Db 362 IWW-----APDSKAVPARFNDTSVDDIRLKNYTN-MENAVISDTKIKYKFGQFQPOF 414
Qy 326 TFMASIMIDAEGRIIDVIDKELIQPEFILLFEGVEYIARAGWTPBCKYASILDRSQR 385
Db 415 DL---FLVNLQNGIILYSINTG-GQKDSILYNG-----KWISDPFRREL-TDRSKI 461
Qy 386 IQVVLISPELFVEEDWNERQQLISVDPSTPLIYESTDIWL-NIHDIHFVQSH 444
Db 462 LDVKYV-----DISSQWL-TYRNTSNLF-----NGWTEKTDIILSIPKPE 503
Qy 445 EEEIE--FIASECKTGPHLYKITSILKESKYKRSSGGLPARSDPKCPKEFIATSGE 502
Db 504 LKRMDDGYIDIHADSGSFHLFYFVF-----AKEPYQLTKGN 542
Qy 503 WEYLGHH--GSNIQVDEVRRLVYFEGTKDPSLEHHLVYVSYV-----NPGEVY 548
Db 543 WEYTGNGIYGVEYED---TIFFANEIGWMSQHLYSILFTSTQNTQNTFQSLQNP----- 594
Qy 549 RLIDRGYHSCGISQHCDFPISKYNSQKNP-----HCYSLYLSSBEDD 552
Db 595 --SDKIDFYDFELSSARAYAISSKRLGPDTPIVAGELTRVLNVAELIHDSIILQLTDE-- 650

```

```

QY 593 PCKTKFEMATILDSAGLPDPYTPPELFSFST---GFTLYGMLYKPHDLPQKPKYPTV 649
Db 651 ---KFKK-----KIKYKULP-ITSYKTMVLDDVDEINYEIRKIPKANINPKKYDIL 696
QY 650 LFTYGGPQVOLVNNRFKGVKFRNLTLASLGVVVVVDNRSGCHRGKLFEGAFKYMGI 709
Db 697 VNIGGSPSSQFTT--KSSLAFEQAVVSGLDVYQIPEPRGTGGGMSFRMARPKYGYW 754
QY 710 EIDDOVE-GLQYLSARYDFIDLDYRGHIGMSYGYLSIMAL-MORSDFRVAIGAPVTL 767
Db 755 EPRITETVTKKFIQNSQIHIDESKAIKMGWSYGFSTLKYVELNDGTFKXAMAVAPVTN 814
QY 768 WIFPDGTERYMGHPDNEQGYIGYVAMQAEKFPSPNNLLHGLFDENVHFAHTSI 827
Db 815 WTLXSVYTERYMQPSENHEGYFEVSTIQNFKSPES-LKRFLFVHGTEFDNVHQLNTER 873
QY 828 LLSFLVRAG-KFYDQIYPERHSTRVPESEGHYTHLHLHLOENLGRIALKVI 882
Db 874 LVDQLNLGLTNYDMHIFPDSHDSIRYHNAQRIYFQKLYWLRDAFAERFDNTEVL 929

RESULT 10
SEPR_HUMAN STANDARD, PRT, 760 AA.
ID Q12884; Q00139; Q99998; Q9UID4;
AC Q12884; Q00139; Q99998; Q9UID4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sepsinase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
DE membrane serine protease) (170-kDa melanoma membrane-bound
DE gelatinase).
GN PAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Butelostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Fibroblast;
RX MEDLINE=94261645; PubMed=7911242;
RA Scanlan M.J., Raj B.K.M., Calvo B., Garin-Chesa P., Sanz-Moncasí M.F.,
RA Healey J.H., Old L.J., Rettig W.J.;
RT "Molecular cloning of fibroblast activation protein alpha, a member of
RT the serine protease family selectively expressed in stromal
RT fibroblasts of epithelial cancers.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:5657-5661(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Melanoma;
RX MEDLINE=97388251; PubMed=9247085;
RA Goldstein L.A., Gherzi G., Pineiro-Sanchez M.L., Salomone M., Yeh Y.,
RA Flessate D., Chen W.-T.;
RT "Molecular cloning of seprase: a serine integral membrane protease
RT from human melanoma.";
RL Biochim. Biophys. Acta 1361:11-19(1997).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 220-229; 461-472 AND
RP 511-518.
RC TISSUE=Melanoma;
RX MEDLINE=97218181; PubMed=9065413;
RA Pineiro-Sanchez M.L., Goldstein L.A., Dotti J., Howard L., Yeh Y.,
RA Chen W.-T.;
RT "Identification of the 170-kDa melanoma membrane-bound gelatinase
RT (seprase) as a serine integral membrane protease.";
RL J. Biol. Chem. 272:7595-7601(1997).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Melanoma;
RX MEDLINE=20112818; PubMed=10644713;
RA Goldstein L.A., Chen W.-T.;
RT "Identification of an alternatively spliced seprase mRNA that encodes
RT a novel intracellular isoform.";

```

```

RL J. Biol. Chem. 275:2554-2559(2000).
RN [5]
RP SEQUENCE OF 192-208; 220-240 AND 510-521.
RX MEDLINE=94327249; PubMed=7519584;
RA Rettig W.J., Su S.L., Fortunato S.R., Scanlan M.J., Raj B.K.M.,
RA Garin-Chesa P., Healey J.H., Old L.J.;
RT "Fibroblast activation protein: purification, epitope mapping and
RT induction by growth factors.";
RL Int. J. Cancer 58:385-392(1994).
CC -1- FUNCTION: May have a role in tissue remodeling during development
CC and wound healing, and may contribute to invasiveness in malignant
CC cancers.
CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC type IV collagen, but not native type I or type IV collagen. Does
CC not cleave laminin, fibronectin, fibrin or casein.
CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC inactive.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Found in cell
CC surface lamellipodia, invadopodia and on shed vesicles.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=L;
CC IsoId=Q12884-1; Sequence=Displayed;
CC Note=Major isoform;
CC Name=2; Synonyms=S, Truncated;
CC IsoId=Q12884-2; Sequence=VSP_005367;
CC -1- TISSUE SPECIFICITY: Fibroblast-specific.
CC -1- INDUCTION: In fibroblasts at times and sites of tissue remodeling
CC during development, tissue repair, and carcinogenesis.
CC -1- PTM: N-glycosylated.
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U09778; AAB49652.1; -;
DR EMBL; U76833; AAC51668.1; -;
DR EMBL; AF007822; AAF21600.1; -;
DR MEROPS; S09_007; -;
DR GeneW; HGNC:3590; PAP.
DR MIM; 600403; -;
DR InterPro; IPR002469; DPP4V N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Prol_endopep_ser.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPP4V N term; 1.
DR Pfam; PF00325; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_S9; 1.
KM Hydroxylase; Protease; Serine protease; Transmembrane; Signal-anchor;
KM Glycoprotein; Alternative splicing.
FT DOMAIN 1 4
FT TRANSMEM 5 25
FT FT 26 760
FT ACT_SITE 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 702 702 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 734 734 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 49 49 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 92 92 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 679 679 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPLIC 1 521 Missing (in isoform 2).
FT FT 207 207 P -> A (IN REF. 2).
FT FT 229 229 K -> T (IN REF. 2).
FT FT 354 354 R -> T (IN REF. 2).
FT CONFLICT 354 354

```



```

Cc 367 EMTPTDDPRMREY-YITWAKATSTKVAWN---LSRQN-----VS-----404
Cc 405 ERQRLIESVPSVPLII--YEETDIWINIHDIHFVFPQSHREELIFASECKTGFRH 462
Cc 405 -----ILTLCDATTCVKKHEDESEANL-----HRQNEPFFVS---KDG-RK 443
Cc 463 LYKITSILK--ESKYKRSGLPAPSDFKCPIKEIATSGEWEVLGRHGSNIQVDEVAR 520
Cc 444 FPFVAIPQGGGGRFYHITVSSSQPNSSNDNIQ---SITSGMDVT---KILSDEKRS 496
Cc 521 LVYFEGTDSPLFHHLYVSVYNPEVTR-----LITRGVSHSCCISQCHDFISKYSN 574
Cc 497 QIYFISTEDLPERRQLYASIV--GSFNRQCLSDLVNCTVFSASFSGADFFLLKCEG 554
Cc 575 QKNPFCVSLYKLSPEDDPTCKE--FWATILDSAGPLPDYPPPIFSFESTGTGLXGM 633
Cc 555 PGVP-TVSVHNTTDKKMFDELTEHNHVKALSDRMKVEYKIB-----TDDYNLIQ 607
Cc 634 LYKPHDLQPKKYPFVLFYGGPQVQVYNNRFGVKYFRLLNTLASLGYVVVINDRSGCH 693
Cc 608 ILKPAFTPDIAHYPLILVVDGTPGSGQVAEKE--AVTWEVWVSSHGAVVVCDSGSGF 665
Cc 694 RGLKEGAFKYMGOIEIDDOVEGLQYLASRYDFIDLRVGHGMSYGYLSIMLMGRS 753
Cc 666 QGTRLLHVRRLGSLERKQMEAVRWL--KEPYIDKTRVAVFGDYGYLSTYLLPAKG 724
Cc 754 D-----IFRVALAGAVTLTMIPTGYTERYMGHPDQNEQGYLGSVMAQEKFPSEPNRL 809
Cc 725 DQGAVFSGSALSTIDTFKLYASAFSERYGLGHLDNRRAHMANVARVSL--EGQGF 782
Cc 810 LLLHGFLENNVFAHTSILLSFVPAKPYDIQIYPOERHSIRVESEHYELHLLHYLQ 869
Cc 783 LVHATADEKIHFORLAEILITQILKGNYSIQIYDESHYFSSAALQHLHRSILGFV 842
Cc 870 E 870
Cc 843 E 843

```

RESULT 12
DPP6_MOUSE STANDARD, PRT, 804 AA.

AC 092218: QSQMW2; 092219, Created
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
GN DPP6 OR DPP-6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=99030650; PubMed=9811881;
RA Hough R.B., Lenzel A., Bedian V., Lo C., Bucan M.;
RT "Rump white inversion in the mouse disrupts dipeptidyl aminopeptidase-
RT like protein 6 and causes dysregulation of kit expression.";
RI Proc. Natl. Acad. Sci. U.S.A. 95:13800-13805(1998).
CC -1- FUNCTION: May be involved in the physiological processes of brain
CC activity. Has no dipeptidyl aminopeptidase activity. The lack of
CC for the serine residue in the proposed catalytic triad (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- SIMILARITY: Belongs to peptidase family S99.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

Cc EMBL; AF092507; AAC97366.1; -
Cc DR EMBL; AF092506; AAC97365.1; -
Cc DR EMBL; AF092505; AAC98381.1; -
Cc DR MEROPS; S09.973; -
Cc DR MGI; 94921; DPP6.
Cc DR InterPro; IPR002469; DPPV_N term.
Cc DR InterPro; IPR001375; Peptidase S9.
Cc DR Pfam; PF00930; DPPV_N term; 1.
Cc DR Pfam; PF00326; Peptidase S9; 1.
Cc KM Transmembrane; Glycoprotein; Signal-anchor.
Cc FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).
Cc FT TRANSMEM 35 55 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
Cc FT DOMAIN 35 55 (POTENTIAL).
Cc FT DOMAIN 56 804 EXTRACELLULAR (POTENTIAL).
Cc FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
Cc FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
Cc FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
Cc FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).
Cc FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
Cc FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
Cc FT CARBOHYD 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).
Cc FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
Cc FT CONFLICT 440 455 MISSING (IN REF. 1; AAC98381).
Cc FT CONFLICT 638 638 O -> P (IN REF. 1; AAC97365).
Cc SQ SEQUENCE 804 AA; 91260 MW; 09CFCE7AD8A7168 CRC64;

```

Query March 9.3%, Score 435; DB 1; Length 804;
Best Local Similarity 23.0%; Pred. No. 6.5e-24;
Matches 178; Conservative 127; Mismatches 322; Indels 148; Gaps 30;

```

Cc 138 LLRERKRIQTVAGIYDHQSGTFLF-----QAGSGIYHVKGDSQPGFTQQLRPN 189
Cc 116 VLIGKKIESRAIRYELSPDXEYVLFYSYNEVPYQSHHTGYVLSKIPHGDPQSLDPE 175
Cc 190 LVETSCPNRMDFLCPADPDWIAFIHSNDIWSINITYREBRRLTYVHNELANNEEDARS 249
Cc 176 VSNAKLQYAGWGPK-----GQQLFIFENNIIYCAHVKAQKARV-----VSTGEGVY 224
Cc 250 AGVATFVLOER-FRYSGYWMCPEKATTPSGCKLRLIYEENDESEVLIHV---TSMVL 305
Cc 225 NGLSDWLYEBEILKSHIAHMSPDG-----TRLAVATINDSRVPLMELPYTGSVY 275
Cc 306 ETRRADSFYKPTGTANPKYTFKMSIMIDAGRILIVIDKELIOPFELFEGVEYIARA 365
Cc 276 PT--VKPYHYKKAASENPISLH-----VIGNGPTHLD--EMMPDDPRMREY-YITMV 325
Cc 366 GWTEBGKA--WILLDRSQRLQIVLSPPLFIVEDDVMERQRLIESVDSVTPPLI- 422
Cc 326 KMATSTVAVVW---LNRQN-----VS-----LITLCDATTCVCTK 359
Cc 423 -YEETDIWINIHDIHFVFPQSHREEL-----EFIFASECKTGFR-HLYKITSILKSK 474
Cc 360 KHEBESAWLH-----RQNEPFFVSQDKRKFVRAIPQGRKGFYHIT--VSSSQ 408
Cc 475 YKRSGLPAPSDFKCPIKEIATSGEWEVLGRHGSNIQVDEVARLVYFEGTDSPLFH 534
Cc 408 PNSSNDNIQ-----SITSGMDVT---KILSDEKRNKIYFISTEDLPERR 451
Cc 535 HLYVSVYNQGEVTRLDRGVSHSCCISQCHDFISKYSNOKNPHCVSLYLSPEDDPT 594
Cc 452 HLVSANTVDPPNRQCL-----SCDLVENCYVASFSHMD--FFLLKCEGP-GVPT 500
Cc 595 C-----KTEFWATILDSAGPLPDYPPPIFSFESTGTGLXGLYVPHD 639
Cc 501 VYHNTTDKRRMPDLANEVQKAIINDRMKPIEYRIEYV-----EDYSLPMQILKPAT 554

```


QY 640 LQPKKXPTVFIYGGPOVOLVNNRFGVKFRLNTLASLGVVVVVINDRSGCHGLKFE 699
 DB 555 FTTDAHPPLLVVGTPOSGVTERFEVT--WETVLVSSHAVVKKCDGRSGFGQSTL 612
 QY 700 GAFKXKQOIEIDDQVEGLQYLASRYDFIDLDVGIHWSYGYLSLMAQMQRSD---I 755
 DB 613 QEVRRRLGFLBKQOMENVRITML--KEGYIDKTRVAVFKDQGGISTYLILPAKGNQGT 671
 QY 756 FRVALAGPVTLMIFYDTGYERYMGHDQNEQGYIGSVAMQAEKFPSEPNRLLLHGF 815
 DB 672 FTCSAISPITDFKLYASAFSEERYLGLGLNRAVEMTKLAHRSAL--EDQOFLIHAT 729
 QY 816 LDENVHAFHTSILSLFVIRACKPYDLOIYPOBRSIRPESGEHYELHLHYLOE 870
 DB 730 ADEKHIFQHTMELITOLIKGKANSLOIYPDESHYFHSVALKQHLRSIGTFVE 784

RESULT 13

DPPE_RAT STANDARD; PRT; 859 AA.
 AC P46101;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
 DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
 DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
 GN DPP6.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
 RC TISSUE=Brain;
 RX MEDLINE=92108018; Pubmed=1729689;
 RA Meda K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
 RT "Differential expression of two distinct forms of mRNA encoding
 RT members of a dipeptidyl aminopeptidase family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
 CC -1- FUNCTION: May be involved in the physiological processes of brain
 CC function. Has no dipeptidyl aminopeptidase activity. The lack of
 CC activity may be due to the substitution of an aspartate residue
 CC for the serine residue in the proposed catalytic triad.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=DPPX-L;
 CC IsoId=P46101-1; Sequence=Displayed;
 CC Name=DPPX-S;
 CC IsoId=P46101-2; Sequence=VSP 005366;
 CC -1- TISSUE SPECIFICITY: DPPX-S IS EXPRESSED IN BRAIN AND SOME
 CC PERIPHERAL TISSUES INCLUDING KIDNEY, OVARY, AND TESTIS; IN
 CC CONTRAST DPPX-L IS EXPRESSED ALMOST EXCLUSIVELY IN BRAIN.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M76426; AAC42061.1; -;
 CC EMBL; M76427; AAC42062.1; -;
 CC MEROPS; S09.973; -;
 CC InterPro; IPR002469; DPPX N term.
 CC InterPro; IPR001375; Peptidase S9.
 CC InterPro; IPR000373; Ser_estrs_site.
 CC Pfam; PF00930; DPPX N term; 1.
 CC Pfam; PF00326; Peptidase S9; 1.
 KW Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.

FT DOMAIN 1 89
 FT TRANSME 90 110
 FT FT
 FT DOMAIN 111 859
 FT CARBOHD 167 167
 FT CARBOHD 168 168
 FT CARBOHD 313 313
 FT CARBOHD 398 398
 FT CARBOHD 465 465
 FT CARBOHD 529 529
 FT CARBOHD 560 560
 FT CARBOHD 807 807
 FT VASPLIC 1 75
 FT FT
 FT FT
 FT FT
 SQ SEQUENCE 859 AA; 97301 MW; CE26856D26BD126B CRC64;
 FT FT
 FT FT
 FT FT

Query Match 9.2%; Score 431; DB 1; Length 859;
 Best Local Similarity 23.0%; Pred. No. 1.4e-23;
 Matches 178; Conservative 129; Mismatches 321; Indels 146; Gaps 30;

QY 138 LLEBRKRGITGSIASYHOSGTF--GAGSGIYHYKDGSGFTQGPLRN 189
 DB 171 VLIGKRTESLRATREISPDKEYALFSYNEPVYQSHGTYVLSKIPIHGDPOSLDPE 230
 QY 190 LVETSCENIRMDPYLCPADPDWIAFIHSNDIMISNITREBRRLTYVNEBLANNEEDARS 249
 DB 231 VSNAKQYAGWGPX-----GQLIFIFENNITVCAHVGKQAIRV-----VSTGEVITY 279
 QY 250 AGVATFVLQER-FRISGYWMCPEAETTPSGKLRILYENDESEVETIIV--TSMVL 305
 DB 280 NGSLDWCVEERILKSHLAHWMSPDG-----TRLAVATINDSRVLMELPYTGSIY 330
 QY 306 ETRADSFYPKGTAPKPKTFKSEIMDAEGRILVIDELLOPEILFEGVEYIARA 365
 DB 331 PT--VKRYHYPKASENPISLH--VIGLNGTHLD--EMPPDPRREY-YITMV 380
 QY 366 GWTPGKXA--WSILDRSQTRLQIVLISPELFIPEVDWEROQLIESVPDVTPLI- 422
 DB 381 KMASTVAATVW--LRAQN-----VS-----ITLCLDATTGCTK 414
 QY 423 -YETDTIMINIHDFHVPQSHBEI-----EFIPASECKTGR-HLYKITSILKESK 474
 DB 415 KHEDESEAMNH-----RQNEPVPFSKDGKRFVRAIPQGRGKFFYHIT--VSSSQ 463
 QY 475 YKSSGGLPAPSDPKCIKEIATISGEMWELGHSNINQVDEVRILVPEGCTDPLEH 534
 DB 464 PNSSNDNIQ-----SITSGMDVIT---ELIYDEKRNKLYFLSTEDLPERR 506
 QY 535 HLIVSYVNGEVTRLNDRGYSHCISQHCDFPISKYSNQKNPFCVSLYKLSSP----- 589
 DB 507 HLVSANVVDPPNRQL-----SCDLVENCITYASAFSINMD--FFLKCEBPQVPTV 556
 QY 590 -EDDPCTKTEF-----WATLDSAGLPDYTPPEITFSFESTGPTLYGMLYKPHD 640
 DB 557 TVNHTTDRKMFDELEANEQYOKAIYDRQMKIEYRKIEV-----EDYSIPMQIKPATF 610
 QY 641 QPKKYPTVFIYGGPOVOLVNNRFGVKFRLNTLASLGVVVVVINDRSGCHGLKFE 700
 DB 611 TDAHPPLLVVGTPOSGVSEFEVT--WETVLVSSHAVVKKCDGRSGFGQSTL 668
 QY 701 AFKXKQOIEIDDQVEGLQYLASRYDFIDLDVGIHWSYGYLSLMAQMQRSD---IF 756
 DB 669 EVRRRLGFLBKQOMENVRITML--KEGYIDKTRVAVFKDQGGISTYLILPAKGNQGT 727
 QY 757 FRVALAGPVTLMIFYDTGYERYMGHDQNEQGYIGSVAMQAEKFPSEPNRLLLHGF 816
 DB 728 TCSSAISPITDFKLYASAFSEERYLGLGLNRAVEMTKLAHRSAL--EDQOFLIHATA 785
 QY 817 LDENVHAFHTSILSLFVIRACKPYDLOIYPOBRSIRPESGEHYELHLHYLOE 870
 DB 786 DEKHIFQHTMELITOLIKGKANSLOIYPDESHYFHSVALKQHLRSIGTFVE 839

CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN NCBI_TaxID=4896;
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21648401; PubMed=11659360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou V., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagele K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Wooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger S., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward V., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Egger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,
RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez V., del Rey F., Bento J.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe.";
RL Nature 415:871-880(2002).
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like vacuoles.
CC -I- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AL162631; CAB83064.1; -;
DR EMBL, Z98596; CAB11208.1; -;
DR GeneDB, SPombe; SPAC14C4.15c; -;
DR InterPro: IPR002469; DPPIV N term.
DR InterPro: IPR001375; Peptidase S9.
DR InterPro: IPR000379; Ser. esterase.
DR Pfam: PF00930; DPPIV N term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
DR Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
KM Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
KW DOMAIN 1 66 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 67 89 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT LUMENAL (POTENTIAL).
FT ACT_SITE 90 853 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 719 719 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 795 795 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 828 828 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 853 AA; 98341 MW; 38450BA50F8304b6 CRC64;
SQ SEQUENCE

[illegible]

Search completed: October 15, 2003, 17:11:26
Job time : 20.9931 secs

Query Match 6.2%; Score 293; DB 1; Length 853;
Best Local Similarity 22.0%; Pred. No. 1.9e-13;
Matches 179; Conservative 110; Mismatches 308; Indels 216; Gaps 38;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:02:16 ; Search time 56.4095 Seconds

(without alignments)
4034.822 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700
Sequence: 1 MAAMETBQLGVETFTADCE.....HLIHYIQENIGSRIALKVI 882

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4700	100.0	882	4	Q9HBM5
2	4696	99.9	882	4	Q8IWM7
3	4528.5	96.4	892	11	Q9DAG6
4	4397.5	93.6	831	4	Q8NEM5
5	3333.5	70.9	632	4	Q96X11
6	2870	61.1	863	4	Q8WMD8
7	2833	60.3	862	11	Q8BWT9
8	2823	60.1	862	11	Q8BVG4
9	2483	52.8	465	4	Q9NMF4
10	2422	51.5	465	4	Q9HBM3
11	2068	44.0	628	4	Q8N3P5
12	1836.5	39.1	360	4	Q9HBM2
13	1645.5	35.0	310	4	Q9HBM4
14	1643.5	35.0	508	4	Q75273
15	1610.5	34.3	439	4	Q9BVR3
16	1599.5	34.0	1042	5	Q9VC20

17	1599.5	34.0	1102	5	Q9VC19	Q9VC19 drosophila
18	1596	34.0	1053	5	Q8IH07	Q8IH07 drosophila
19	1175	25.0	312	4	Q96NTR	Q96NTR homo sapien
20	1049	22.3	432	4	Q75868	Q75868 homo sapien
21	924.5	19.7	927	5	Q965K3	Q965K3 caenorhabdi
22	922.5	19.6	931	5	Q44987	Q44987 shewanella
23	846	18.0	763	16	Q8EAB7	Q8EAB7 shewanella
24	842	17.9	738	16	Q9A6E0	Q9A6E0 caulobacter
25	837.5	17.8	746	10	Q9FNP6	Q9FNP6 arbidopsis
26	754	16.0	741	2	P95782	P95782 xanthomonas
27	751.5	16.0	751	16	Q8P3V8	Q8P3V8 xanthomonas
28	744.5	15.8	757	16	Q8EPD7	Q8EPD7 xanthomonas
29	727	15.5	552	10	Q8GUJ7	Q8GUJ7 arabidopsis
30	664	14.1	167	4	Q8N2J7	Q8N2J7 homo sapien
31	620	13.2	711	2	Q47900	Q47900 flavobacter
32	582.5	12.4	771	3	Q42812	Q42812 aspergillus
33	572.5	12.2	723	2	Q66223	Q66223 porphyromon
34	563.5	12.0	723	2	Q31048	Q31048 porphyromon
35	553.5	11.8	901	3	Q96VT7	Q96VT7 aspergillus
36	546	11.6	730	2	Q93JY4	Q93JY4 prevotella
37	539.5	11.5	765	3	Q14425	Q14425 aspergillus
38	537.5	11.4	748	13	P70092	P70092 xenopus lae
39	484.5	10.3	761	11	Q8R492	Q8R492 ratcus norv
40	474	10.1	707	16	Q9F348	Q9F348 streptomyce
41	468.5	10.0	814	16	Q8G6O6	Q8G6O6 bifidobacte
42	463.5	9.9	935	5	Q9VMB4	Q9VMB4 drosophila
43	462	9.8	796	4	Q8N6O8	Q8N6O8 homo sapien
44	461	9.8	711	4	Q9P236	Q9P236 homo sapien
45	461	9.8	802	5	Q9VTH1	Q9VTH1 drosophila

ALIGNMENTS

RESULT 1
Q9HBM5 PRELIMINARY; PRT; 882 AA.

101-MAR-2001 (TREMBLrel. 16, Created)
01-MAR-2001 (TREMBLrel. 16, Last sequence update)
01-MAR-2003 (TREMBLrel. 23, Last annotation update)
Dipeptidyl peptidase 8.
DPP8.
Homo sapiens (Human).
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=20467194; PubMed=11012666;
Abbot C.A., Yu D.M.T., Woollett E., Sutherland G.R., McCaughan G.W.,
Gorell M.D.;
"Cloning, expression and chromosomal localization of a novel human
dipeptidyl peptidase (DPP) IV homolog, DPP8.";
Eur. J. Biochem. 267:6140-6150(2000).
EMBL; AF221634; AAG29766.1; -.
MEROPS; S09.018; -.
InterPro; IPR002469; DPPIV N term.
InterPro; IPR001375; Peptidase_S9.
InterPro; IPR00379; Ser_estr_site.
Pfam; PF00930; DPPIV N term; 1.
Pfam; PF00326; Peptidase_S9; 1.
SEQUENCE 882 AA; 101421 MW; AD801C302DB4652B CRC64;

Query Match 100.0%; Score 4700; DB 4; Length 882;

Best Local Similarity 100.0%; Pred.No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAMETBQLGVETFTADCEINISODPKLEPFYVERYSQKLIADRKHYGVM 60
DB 1 MAAMETBQLGVETFTADCEINISODPKLEPFYVERYSQKLIADRKHYGVM 60
QY 61 AKAPEHMFVKRNDPDGPHSDRIYYLAWMSGENRENTLFYSELPKTIINRAAVLMLSWKPLL 120

```

Db      61 AKAHPDFEVRKNDPDGPHSDRIYYLAAMSGENRENTLFYSEIPTKINRAAVLMSWKPL 120
Qy      121 DLFOATLDYGYMSSEEBELRERKRIIGTVGASVDYHOGSGTFLQAGSGIYHVDGGQG 180
Db      121 DLFOATLDYGYMSSEEBELRERKRIIGTVGASVDYHOGSGTFLQAGSGIYHVDGGQG 180
Qy      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPMIAFIHNSNDIWSNIVTREERLTYVHNE 240
Db      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPMIAFIHNSNDIWSNIVTREERLTYVHNE 240
Qy      241 ANNEEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEI 300
Db      241 ANNEEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEI 300
Qy      301 TSPMLERRADSFRPYKGTANPKVTFKMSIIMDAEGRIIDVIDKELIQFELLFEGVE 360
Db      301 TSPMLERRADSFRPYKGTANPKVTFKMSIIMDAEGRIIDVIDKELIQFELLFEGVE 360
Qy      361 YIARAGTPEGKYAMSILLDRSQTRLOIVLISPFLFIVEDDVMERORLIESVDSVTP 420
Db      361 YIARAGTPEGKYAMSILLDRSQTRLOIVLISPFLFIVEDDVMERORLIESVDSVTP 420
Qy      421 IYEBETDWINHNDIHVPQSHHEEIEFIASECTGFRHLKYITSILKESKYKSSG 480
Db      421 IYEBETDWINHNDIHVPQSHHEEIEFIASECTGFRHLKYITSILKESKYKSSG 480
Qy      481 GLPAPSDPKCPKEKEIAITSGEWEVLGRHGSNIQVDEVRLVYEGTKDSPLHHLVVS 540
Db      481 GLPAPSDPKCPKEKEIAITSGEWEVLGRHGSNIQVDEVRLVYEGTKDSPLHHLVVS 540
Qy      541 YVNGEVTRLTDRGYSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTEF 600
Db      541 YVNGEVTRLTDRGYSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTEF 600
Qy      601 WATILDSAGPLDPTTPPEIISFESTTGFTLYGMLYKPHDLOPKKYPTVLFIYGGPOVL 660
Db      601 WATILDSAGPLDPTTPPEIISFESTTGFTLYGMLYKPHDLOPKKYPTVLFIYGGPOVL 660
Qy      661 VNNRFKGVKVFRLNTLASLGYVVVVDNRGSGCHRGKFEKGFYKMGQIEIDQVEGLQY 720
Db      661 VNNRFKGVKVFRLNTLASLGYVVVVDNRGSGCHRGKFEKGFYKMGQIEIDQVEGLQY 720
Qy      721 LASRYDFIDLDRAVIGHMSYGYLSLMAIMORSDFRVALAGAVTLMIFDGTGTERYM 780
Db      721 LASRYDFIDLDRAVIGHMSYGYLSLMAIMORSDFRVALAGAVTLMIFDGTGTERYM 780
Qy      781 GHPDQNEQGYLLGSVMAQAEKFPSEPNRLILHGFLEBNVHFAHTSILLSFLVAGKPYD 840
Db      781 GHPDQNEQGYLLGSVMAQAEKFPSEPNRLILHGFLEBNVHFAHTSILLSFLVAGKPYD 840
Qy      841 LQIYPERHSIRVPSGEHVELHLHYLOENLSRIAALKVY 882
Db      841 LQIYPERHSIRVPSGEHVELHLHYLOENLSRIAALKVY 882

```

RESULT 2

```

ID      081WG7      PRELIMINARY;      PRT;      882 AA.
AC      081WG7;
DT      01-MAR-2003 (Tremblrel. 23, Created)
DT      01-MAR-2003 (Tremblrel. 23, last sequence update)
DE      01-MAR-2003 (Tremblrel. 23, last annotation update)
DE      Similar to dipeptidylpeptidase 8.
OS      Homo sapiens (Human).
OC      Bakayola; Metacoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Testis;
RA      Strausberg R.;
RL      Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

```

```

DR      EMBL, BC040203; AAH40203.1; -.
SQ      SEQUENCE      882 AA; 101391 MM; 88C76A5BCE707F9 CRC64;
Query Match      99.9%; Score 4696; DB 4; Length 882;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 881; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy      1 MAAMEEQGLVEIFETADCEENIESODRPKLEFFYERYWSQJLKLADTRKYHGM 60
Db      1 MAAMEEQGLVEIFETADCEENIESODRPKLEFFYERYWSQJLKLADTRKYHGM 60
Qy      61 AKAHPDFEVRKNDPDGPHSDRIYYLAAMSGENRENTLFYSEIPTKINRAAVLMSWKPL 120
Db      61 AKAHPDFEVRKNDPDGPHSDRIYYLAAMSGENRENTLFYSEIPTKINRAAVLMSWKPL 120
Qy      121 DLFOATLDYGYMSSEEBELRERKRIIGTVGASVDYHOGSGTFLQAGSGIYHVDGGQG 180
Db      121 DLFOATLDYGYMSSEEBELRERKRIIGTVGASVDYHOGSGTFLQAGSGIYHVDGGQG 180
Qy      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPMIAFIHNSNDIWSNIVTREERLTYVHNE 240
Db      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPMIAFIHNSNDIWSNIVTREERLTYVHNE 240
Qy      241 ANNEEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEI 300
Db      241 ANNEEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEI 300
Qy      301 TSPMLERRADSFRPYKGTANPKVTFKMSIIMDAEGRIIDVIDKELIQFELLFEGVE 360
Db      301 TSPMLERRADSFRPYKGTANPKVTFKMSIIMDAEGRIIDVIDKELIQFELLFEGVE 360
Qy      361 YIARAGTPEGKYAMSILLDRSQTRLOIVLISPFLFIVEDDVMERORLIESVDSVTP 420
Db      361 YIARAGTPEGKYAMSILLDRSQTRLOIVLISPFLFIVEDDVMERORLIESVDSVTP 420
Qy      421 IYEBETDWINHNDIHVPQSHHEEIEFIASECTGFRHLKYITSILKESKYKSSG 480
Db      421 IYEBETDWINHNDIHVPQSHHEEIEFIASECTGFRHLKYITSILKESKYKSSG 480
Qy      481 GLPAPSDPKCPKEKEIAITSGEWEVLGRHGSNIQVDEVRLVYEGTKDSPLHHLVVS 540
Db      481 GLPAPSDPKCPKEKEIAITSGEWEVLGRHGSNIQVDEVRLVYEGTKDSPLHHLVVS 540
Qy      541 YVNGEVTRLTDRGYSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTEF 600
Db      541 YVNGEVTRLTDRGYSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTEF 600
Qy      601 WATILDSAGPLDPTTPPEIISFESTTGFTLYGMLYKPHDLOPKKYPTVLFIYGGPOVL 660
Db      601 WATILDSAGPLDPTTPPEIISFESTTGFTLYGMLYKPHDLOPKKYPTVLFIYGGPOVL 660
Qy      661 VNNRFKGVKVFRLNTLASLGYVVVVDNRGSGCHRGKFEKGFYKMGQIEIDQVEGLQY 720
Db      661 VNNRFKGVKVFRLNTLASLGYVVVVDNRGSGCHRGKFEKGFYKMGQIEIDQVEGLQY 720
Qy      721 LASRYDFIDLDRAVIGHMSYGYLSLMAIMORSDFRVALAGAVTLMIFDGTGTERYM 780
Db      721 LASRYDFIDLDRAVIGHMSYGYLSLMAIMORSDFRVALAGAVTLMIFDGTGTERYM 780
Qy      781 GHPDQNEQGYLLGSVMAQAEKFPSEPNRLILHGFLEBNVHFAHTSILLSFLVAGKPYD 840
Db      781 GHPDQNEQGYLLGSVMAQAEKFPSEPNRLILHGFLEBNVHFAHTSILLSFLVAGKPYD 840
Qy      841 LQIYPERHSIRVPSGEHVELHLHYLOENLSRIAALKVY 882
Db      841 LQIYPERHSIRVPSGEHVELHLHYLOENLSRIAALKVY 882

```

RESULT 3

```

ID      09D4G6      PRELIMINARY;      PRT;      892 AA.
AC      09D4G6;
DT      01-JUN-2001 (Tremblrel. 17, Created)

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Adult male testis cDNA, RIKEN full-length enriched library,
 DE clone:493244F09 product:DIPEPTIDYL PEPTIDASE 8 homolog.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA Aichi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arawaka T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoaka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Saeki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA MEDLINE=22354683; PubMed=1246851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.,
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA MEDLINE=2049374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Sumi N., Iehi Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashinaga K.,
 RA Fujitake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe K.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsunaka S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 DR EMBL, A015646; BAB30295.2; --
 SQ SEQUENCE 892 AA; 102284 MW; FDE0DBDDC4CA346 CRC64;

Query Match 96.4%; Score 4528.5; DB 11; Length 892;

Best Local Similarity 95.7%; Pred. No. 0;
 Matches 845; Conservative 18; Mismatches 19; Indels 1; Gaps 1;
 QY 1 MAAMETEQGVLEIFETADCE-NISQDCKPEPFVVEYSSQKLLADPRKHGM 59
 DB 10 MAAMETEQGVLEIFETAECEBNGESQDPEKPEPFVVEYSSQKLLADPRKHGM 69
 QY 60 MACAPDFMFVKNDPDSRIYYLAMSNGENRENTLFYSELPKTIINRAAVLMLSWKPL 119
 DB 70 MACAPDFMFVKNDPDRPHSDRAVYLLAMSGENRENTLFSELPKTIINRAAVLMLSWKPL 129
 QY 120 LDFPQATLDYGMTSREBELREKRIGTGVIAVDHQSGTFLPQAGSGIYHYKDGQP 179
 DB 130 LDFPQATLDYGMTSREBELREKRIGTGVIAVDHPGSGTFLPQAGSGIYHYKDGSPH 189
 QY 180 GFTQOPRLPVLTSCGNIRMDKLCAPDDWIAFHSNIDWISNITREERRLTYYNE 239
 DB 190 GFTQOPRLPVLTSCGNIRMDKLCAPDDWIAFHSNIDWISNITREERRLTYYNE 249
 QY 240 LANMEDARSAGVATFVLOREDFRYSGYMWCPXAEPTPSGKILRIYEENDESEVEIYH 299
 DB 250 LANMEDARSAGVATFVLOREDFRYSGYMWCPXAEPTPSGKILRIYEENDESEVEIYH 309
 QY 300 VTSPMLSTRADSPRYPKTGTANPKYTFKKSSEIMIDAGRIIDYIDKELIQPEFIIEGV 359
 DB 310 VTSPMLSTRADSPRYPKTGTANPKYTFKKSSEIVVDAAGIIVYIDKELVQPEFIIEGV 369
 QY 360 EYIARAGWTPREGVYASILLDRSOTRLQYILSPELFIPEDDVMERQRLIESVDSVTP 419
 DB 370 EYIARAGWTPREGVYASILLDRSOTRLQYILSPELFIPEDDVMERQRLIESVDSVTP 429
 QY 420 LIYEETDWINIHIDIFHVPOSHBELEFIFASECKTFRHLKYTSLIKESKYRSS 479
 DB 430 LIYEETDWINIHIDIFHVPOSHBELEFIFASECKTFRHLKYTSLIKESKYRSS 489
 QY 480 GGIAPASDFKCPKEEALITSGEWYLGRRGSNIQVDEVRLTYFEGTKOSPLEHLLYV 539
 DB 490 GGIAPASDFKCPKEEALITSGEWYLGRRGSNIQVDEVRLTYFEGTKOSPLEHLLYV 549
 QY 540 SYVNGEVRRLTRRGVSHSCCISQHCDFEISXSNQNCPCVSLYKLSPEDDPTCKTXE 599
 DB 550 SYVNGEVRRLTRRGVSHSCCISQHCDFEISXSNQNCPCVSLYKLSPEDDPTCKTXE 609
 QY 600 FMATILDSAGPLDYTPPELIFSPESSTGFTLYGMLYKPHDLOPKKYPVLFYIGSPQV 659
 DB 610 FMATILDSAGPLDYTPPELIFSPESSTGFTLYGMLYKPHDLOPKKYPVLFYIGSPQV 669
 QY 660 LVNRRFKGYKYPFLNLTASLGIVYVYVITDNRGSGHRGLKFGAARKYKKGQIEIDQVGLQ 719
 DB 670 LVNRRFKGYKYPFLNLTASLGIVYVYVITDNRGSGHRGLKFGAARKYKKGQIEIDQVGLQ 729
 QY 720 YLASRYDFIDLDRVGIHGSYGGYLSLMLMORSDFRVAIAAPVTLMTFYDTGYTERY 779
 DB 730 YLASRYDFIDLDRVGIHGSYGGYLSLMLMORSDFRVAIAAPVTLMTFYDTGYTERY 789
 QY 780 MGHPDONEQGYLYGSVAMQAEKPESEPNRLLHGFLENVHFAHTSILLSPYRAGKPY 839
 DB 790 MGHPDONEQGYLYGSVAMQAEKPESEPNRLLHGFLENVHFAHTSILLSPYRAGKPY 849
 QY 840 DLQIYPOBHSIRVPESGHEHLHLHYOENIGSILAAKVI 882
 DB 850 DLQIYPOBHSIRVPESGHEHLHLHYOENIGSILAAKVI 892
 RESULT 4
 O8NEMS PRELIMINARY; PRT; 831 AA.
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Similar to dipeptidyl peptidase 8.
 OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Strausberg R.;
 RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC030688; AAH30688.1; -
 SQ SEQUENCE 831 AA; 95527 MW; 0B2A13A2FE70CB2 CRC64;

Query Match 93.6%; Score 4397.5; DB 4; Length 831;
 Best Local Similarity 94.2%; Pred. No. 0;
 Matches 831; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

QY 1 MAAMETEQGLVETADCEENIESODRPLEPFYERYSMSQKLLADTRKHYGM 60
 Db 1 MAAMETEQGLVETADCEENIESODRPLEPFYERYSMSQKLLADTRKHYGM 60
 QY 61 AKAPHEMFVYKNDPDGPHSDRIYLLAMSGENRENTLFYSEIPTINRAAVLMSKPL 120
 Db 61 AKAPHEMFVYKNDPDGPHSDRIYLLAMSGENRENTLFYSEIPTINRAAVLMSKPL 120
 QY 121 DLPAATLDYGMYSSEELERKIGTVGTAIYNHOGSGTFLFOAGSGIYHVDGPG 180
 Db 121 DLPAATLDYGMYSSEELERKIGTVGTAIYNHOGSGTFLFOAGSGIYHVDGPG 180
 QY 181 FTQGPLRPNIYETSCPNIRMDPKLCPADPWIAFIHNSDIWISNITREERLLTYANEL 240
 Db 181 FTQGPLRPNIYETSCPNIRMDPKLCPADPWIAFIHNSDIWISNITREERLLTYANEL 240
 QY 241 ANMEDARSAGVATFVIOEEDRISGYWCPKATTPSGKILRIYENDESEVEIIVH 300
 Db 241 ANMEDARSAGVATFVIOEEDRISGYWCPKATTPSGKILRIYENDESEVEIIVH 300
 QY 301 TSPMLERRRDSFPYKGTANPKYTFMSEIMIDAGRIIDVDKELQPELLEFGVE 360
 Db 301 TSPMLERRRDSFPYKGTANPKYTFMSEIMIDAGRIIDVDKELQPELLEFGVE 360
 QY 361 YIARAGWTPGKYAMSILDRSQRLQIVLISPFLIPVEDDWERORLIESVDSVTP 420
 Db 361 YIARAGWTPGKYAMSILDRSQRLQIVLISPFLIPVEDDWERORLIESVDSVTP 420
 QY 421 IYETTDIWINIHDIHVPPOSHHEEIEFIASECTGFRHLKYTTSIIKESKYKSSG 480
 Db 421 IYETTDIWINIHDIHVPPOSHHEEIEFIASECTGFRHLKYTTSIIKESKYKSSG 480
 QY 481 GLPAPSPFCPIKEELAITSGEWEVLGRHGSNIQVDEVRILYVEGTKDSPLHHLVYS 540
 Db 481 GLPAPSPFCPIKEELAITSGEWEVLGRHGSNIQVDEVRILYVEGTKDSPLHHLVYS 540
 QY 541 YVNPGEVTRLDRGYSHSCISQCHDFIFISKYSNOKNPHCVSLYKLSPPEDPTCKTKEF 600
 Db 541 YVNPGEVTRLDRGYSHSCISQCHDFIFISKYSNOKNPHCVSLYKLSPPEDPTCKTKEF 600
 QY 601 MATLLDSAGLDPYTPPEIISFESTTFTLYGMLYKHDIQPGKYTVVFIYGGPOVL 660
 Db 601 MATLLDSAGLDPYTPPEIISFESTTFTLYGMLYKHDIQPGKYTVVFIYGGPOVL 660
 QY 661 VNNRFKGVKFFRLNTLASLGIVVVIDNRSGCHRGKFEKAFKYMGOIIDDQVEGLQ 720
 Db 661 VNNRFKGVKFFRLNTLASLGIVVVIDNRSGCHRGKFEKAFKYMGOIIDDQVEGLQ 720
 QY 721 LASRYDFIDLVGIGHWSYGYISLALMORSIDFVAIAGAPVTLIMIFYDTGYTERY 780
 Db 721 LASRYDFIDLVGIGHWSYGYISLALMORSIDFVAIAGAPVTLIMIFYDTGYTERY 780
 QY 781 GHDPDNOGGYLLGSVAMQAEKPPSEPRLLLLHGLDENHFAHTSILSLVARAGPYD 840
 Db 781 GHDPDNOGGYLLGSVAMQAEKPPSEPRLLLLHGLDENHFAHTSILSLVARAGPYD 840
 QY 841 LQIYPERHSIRVPESGEHYELHLHYQENLGSRIALAKVI 882
 Db 841 LQIYPERHSIRVPESGEHYELHLHYQENLGSRIALAKVI 882

Db 790 LQIYPERHSIRVPESGEHYELHLHYQENLGSRIALAKVI 831

RESULT 5

Q96JXI PRELIMINARY; PRT; 632 AA.
 AC Q96JXI;
 DT 01-DEC-2001 (TREMUREL 19, Created)
 DT 01-DEC-2001 (TREMUREL 19, Last sequence update)
 DT 01-OCT-2002 (TREMUREL 22, Last annotation update)
 DE Hypothetical protein FLJ14920.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
 RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Takiguchi S.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
 RA Niimura K., Iwayanagi T.;
 RT "NBD human cDNA sequencing project."
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK027826; BAB55395.1; -
 DR MEROPS; S09_018; -
 KW Hypothetical protein.
 SQ SEQUENCE 632 AA; 72639 MW; 9BDF598B06985AA4 CRC64;

Query Match 70.9%; Score 3333.5; DB 4; Length 632;
 Best Local Similarity 92.4%; Pred. No. 1.9e-250;
 Matches 631; Conservative 0; Mismatches 1; Indels 51; Gaps 1;

QY 200 MDKLCPADPDWIAFIHNSDIWISNITVREERLLTYVHNEIAMMEDARSAGVATFVLOE 259
 Db 1 MDKLCPADPDWIAFIHNSDIWISNITVREERLLTYVHNEIAMMEDARSAGVATFVLOE 259
 QY 260 EFRYSGYWWCPKATTPSGKILRIYENDESEVEIIVHTSPMLERRRDSFPYKGT 319
 Db 61 EFRYSGYWWCPKATTPSGKILRIYENDESEVEIIVHTSPMLERRRDSFPYKGT 319
 QY 320 TANPKYTFKSEIMIDAGRIIDVDKELQPELIFEGEYIARAGWTPGKYAMSIL 379
 Db 121 TANPKYTFKSEIMIDAGRIIDVDKELQPELIFEGEYIARAGWTPGKYAMSIL 379
 QY 380 DRSQRLQIVLISPFLIPVEDDWERORLIESVDSVTPLIYETTDIWINIHDIHVP 439
 Db 181 DRSQRLQIVLISPFLIPVEDDWERORLIESVDSVTPLIYETTDIWINIHDIHVP 439
 QY 440 FPGSHHEEIEFIASECTGFRHLKYTTSIIKESKYKSSGCPAPSPFCPIKEELAIT 499
 Db 241 FPGSHHEEIEFIASECTGFRHLKYTTSIIKESKYKSSGCPAPSPFCPIKEELAIT 499
 QY 500 SCGEWEVLGRHGSNIQVDEVRILYVEGTKDSPLHHLVYSVNPGEVTRLDRGYSHSC 559
 Db 301 SCGEWEVLGRHGSNIQVDEVRILYVEGTKDSPLHHLVYSVNPGEVTRLDRGYSHSC 559
 QY 560 CISOCHDFIFISKYSNOKNPHCVSLYKLSPPEDPTCKTKEFMAITILDSAGPLDPYTPPEI 619
 Db 361 CISOCHDFIFISKYSNOKNPHCVSLYKLSPPEDPTCKTKEFMAITILDSAGPLDPYTPPEI 619
 QY 620 FPFESTTFTLYGMLYKHDIQPGKYTVVFIYGGPOVLVNNRFKGVKFFRLNTLASL 679
 Db 421 FPFESTTFTLYGMLYKHDIQPGKYTVVFIYGGPOVLVNNRFKGVKFFRLNTLASL 679
 QY 680 GYVVVVVIDNRSGCHRGKFEKAFKYMGOIIDDQVEGLQYLASRYDFIDLVGIGHWS 739
 Db 481 GYVVVVVIDNRSGCHRGKFEKAFKYMGOIIDDQVEGLQYLASRYDFIDLVGIGHWS 739
 QY 740 YGYISLALMORSIDFVAIAGAPVTLIMIFYDTGYTERYMGHPDNOGGYLLGSVAMQ 799
 Db 740 YGYISLALMORSIDFVAIAGAPVTLIMIFYDTGYTERYMGHPDNOGGYLLGSVAMQ 799


```
Db 508 -----VALAGAPVTLMTFYDTGTERYMGHPDQNGYLLSVAQA 549
Qy 800 EKFPSEPNRLLLHGFIDENVAFAHTSLLSFLVAGKPYDLOIYPOERHSIRVDSGEH 859
Db 550 EKFPSEPNRLLLHGFIDENVAFAHTSLLSFLVAGKPYDLOIYPOERHSIRVDSGEH 609
Qy 860 YELHLHYLOENLGSRIALAVI 882
Db 610 YELHLHYLOENLGSRIALAVI 632

RESULT 6
Q8WXD8 PRELIMINARY; PRT; 863 AA.
ID Q8WXD8
AC Q8WXD8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Dipeptidyl peptidase 9.
GN DPP9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid=9606;
RN [1];
RP SEQUENCE FROM N.A.
RA Olsen C., Maglmann N.;
RT "Identification and characterization of a novel member of the
RT dipeptidyl peptidase IV-related family";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [2];
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RT Tissue=Skin;
RC Tissue=Skin;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF452102; AAL47179.1; -.
DR EMBL; BC037948; AAH37948.1; -.
DR MEROPS; S09.019; -.
DR InterPro; IPR002469; DPPIV N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser esterase_site.
DR Pfam; PF00930; DPPIV N term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
SQ SEQUENCE 863 AA; 98263 MW; 40FEDB78E26CDED5 CRC64;

Query Match 61.1%; Score 2870; DB 4; Length 863;
Best Local Similarity 61.5%; Pred. No. 3.4e-214;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;
```

```
Qy 394 ELFIPEVDVNEROLLESVDSYTPLIYEETDIWINIHDPHPQSH-EETIEFI 452
Db 384 ALFIPTSENEROLASARAARVNAVYVVEEYVNWIMVHDIYFPFQSGGBELCTL 443
Qy 453 ASECKTGRHLYKITSILKESKTKYRSSGGLPAPSDFKPIKEEIALISGEWVYGRHSN 512
Db 444 ANECKTGFCHLYKTAVLVKSGYDMSERFSGDEFKPIKEEIALISGEWVYGRHSK 503
Qy 513 IQDEVRILYFECKTSPLEHNLVYVYVNWVGTPLTDVSGHSCCISGHCPEFIKY 572
Db 504 IWNVEETLVYFQGTCTPLEHNLVYVYVNWVGTPLTDVSGHSCCISGHCPEFIKY 563
Qy 573 SNQKRPCHVSLYKSSDEDPCTCKEFATILDSAGLPDYTPPEISFSTGFTLYG 632
Db 564 SSVSTPCVHYKTLGSDDDDLHQPFPWASMEASCPDYVPEIHFHTRSVRLYG 623
Qy 633 MLYRPHLOPKKPYTLFTYGGQVOLLVNNRFGVYKFRNTLASLGYVAVVDNRSGC 692
Db 624 MYRPHALOPKPKPYTLFTYGGQVOLLVNNRFGVYKFRNTLASLGYVAVVDNRSGC 683
Qy 693 HRGLKFGAPFYKMGQIEIDQVGLASRYDFIDLRGIGHGWSYGYLSMALMOR 752
Db 684 QRGKLFEGALKNQGVIEIDQVGLQPVAEKGFIDLSRVALHGWSTGYGLSLMGLTHK 743
Qy 753 SDIFRVAIAGAPVTLMTFYDTGTERYMGHPDQNGYLLSVAQA EKFPSEPNRLLL 812
Db 744 PQVFKVAILAGAPVTLMTFYDTGTERYMGHPDQNGYLLSVAQA EKFPSEPNRLLL 803
Qy 813 HGFIDENVAFAHTSLLSFLVAGKPYDLOIYPOERHSIRVDSGEH YELHLHYLOENL 872
Db 804 HGFIDENVAFAHTSLLSFLVAGKPYDLOIYPOERHSIRVDSGEH YELHLHYLOENL 863

RESULT 7
Q8BWT9 PRELIMINARY; PRT; 862 AA.
ID Q8BWT9
AC Q8BWT9;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Dipeptidyl peptidase 9 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1];
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; Tissue=Liver;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
DR EMBL; AK050021; BAC34034.1; -.
SQ SEQUENCE 862 AA; 98001 MW; B1D56E824A834B8 CRC64;

Query Match 60.3%; Score 2833; DB 11; Length 862;
Best Local Similarity 60.8%; Pred. No. 2.6e-211;
Matches 511; Conservative 134; Mismatches 193; Indels 2; Gaps 2;
```

QY	35	FFYERYNSQLKLLADTRKHYGYMAKAPHDNFVYKRPNDGCHSPRIYYLAMSNGRE	94
	23	FCVKGHSVDGRSLIHSKRSSGLIVSKAPHDFQVOKPSPESGPHSHRIYYLGMVQGRE	82
QY	95	NLFSEIPEPKINAAVYLMWKRLDLPFOATLDYGYSSHEELLRRKXIIGTGISYD	154
	83	NSLYSEIPEPKYRKREALILSMKQMLHFQATPHHGYSSHEELLRRKXIGFGLISYD	142
QY	155	YHOGSGTFLFOAGSGIYHVXDGGPQGFOTQOPLREBNILETSCPNIRMPKLCPADPWIAF	214
	143	FHSSGGLFLPQANSILFHCNRDGGKGMVSWMKLEIKTQCSSGRMOPKICPADPAFISF	202
QY	215	IHSNDIWINI VTRBERLLYVHNELANMBEDARSAGVATVLOEEDFRYSGWVCKPAE	274
	203	INNSDILWVANITEEERRLTFCHOGSGAGVLDNPKSAGVATFVIOEEDFRFGWCWCPAS	262
QY	275	TTSPGG-KILRIYKLENDSEVEIYIHTSPMLFTRADSPRYPKTGANPVTFKSEIM	333
	263	WESSEGLTILKILVEBDESEVEIYIHPSPLEERKTDSTRYPTGSKNPIALUKIAELQ	322
QY	334	IDABGRILIDVIDKLEIOPFELIEFGVYIYIAGVTEPEKAWMSILLRSOTRLQIVLISP	393
	323	TDHGKIVSSCEKELVQPFSLFPEKVEYIYIAGVTRGKXAMAMFLVRPQORQIVLILP	382
QY	394	ELFIFVEDDWERKRLLESVPDSTPLIYIETTDIYINIHDFHVPQSH-EEIEIFIF	452
	383	ALFIPAYESEKORQAARVAPKVVQPVYIEEYVNIWVINHDIHFPPPOAGMODFELR	442
QY	453	ASECKTGPFRHLYKTSILKESKYRSSGGLPAPDPCPIKEEIALTSGMEVLYGRGNS	512
	443	ANECKTGPCHLYRTVLELTKTDWTEPLSPTEDEFCPIKEEVALTSGMEVLYSRGSK	502
QY	513	IQVDEVARLYFEGTKOSPLEHHLVYVSYNVPGEVTRKLDRGYSHSCISQACDFFISKY	572
	503	IWNVEQKLVYFQGTQOTPLEHHLVYVSYSAEIVALTLLTGLFSGHSCMSQSDMFVSHY	562
QY	573	SNQKNPFCVSLYKLSSEDPOTCKYEFATITLDSAGPLPDYTPPEIYESSTGFTLYG	632
	563	SSVSTPCCVHYKLSGDDDDPLKPOPRFMASSMEANCPDYPPELIFHPTRADVQLYG	622
QY	633	MLYXPHLOPGKKPYTLFYGGGPOVLYNNRFGYVYFPLNTLASIGYVYVYINDRGSC	692
	623	MIYXPHLOGRKHPIVLFYGGGPOVLYNNRFGYVYFPLNTLASIGYVYVYINDRGSC	682
QY	693	HRLGPEGAARYKNGOIEIDQVEGLQYLASRYDFIDLDRVGTHGMSYGGYSILAMOR	752
	683	QRLGHFEAGLKNQGCQVEIEDQVEGLQYLAEXKGFILDSVALIHGMSYGGYSILMGLIKH	742
QY	753	SDIRVVALAAGPYTLMFFYDTGTYTERMYGPHDQNEGGYIYGYAMQLEKFPSPENRLLIL	812
	743	PQYFKVAILAAGPYVMMAYIDQYTERMYDVENNNKQGYEAGSYALHVEKLPENRNLIL	802
QY	813	HGFIDEVNEFAHPSILSLPVRAGKRPDIQIYQBSHSTRVPSGGHYELHLYAOENL	872
	803	HGFIDEVNEFAHPSILSLPVRAGKRPDIQIYENESHSTRCSGGEHYETVTLHPLQENL	862
RESULT 9			
ID	Q9NXP4	PRELIMINARY;	PRT; 469 AA.
AC	Q9NXP4;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)		
DE	Hypothetical protein FLJ20283.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euteheria; Primates; Catartini; Homiidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RA	Kawakami T., Nobuchi S., Itoh T., Shigeta K., Senba T., Matsumura K.		

RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
 RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isegaki T., Sugano S.,
 RT "NEDO human cDNA sequencing project";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK000230; BAA91059.1; -
 DR MEROPS; S09.018; -
 KW Hypothetical protein.

SO SEQUENCE 469 AA; 54367 MW; 088ED0B1E46C11F CRC64;

Query Match 52.8%; Score 2483; DB 4; Length 469;
 Best Local Similarity 100.0%; Pred. No. 1,7e-184;
 Matches 469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAMETEQGLVEIFETADCEENIESQDPKLEPPYERYSWQLKTLADTRYKGYMM 60
 DB 1 MAAMETEQGLVEIFETADCEENIESQDPKLEPPYERYSWQLKTLADTRYKGYMM 60
 QY 61 AKAPHDPMFYKRNPDGPHSDRIYLLAMSGENRENTLFYSRIPTINRAALMSWPL 120
 DB 61 AKAPHDPMFYKRNPDGPHSDRIYLLAMSGENRENTLFYSRIPTINRAALMSWPL 120
 QY 121 DLFOATLDYGYMSSEBELRERKRIQTVGLASYDHOSSGTFLEQAGSGIYHVKGPG 180
 DB 121 DLFOATLDYGYMSSEBELRERKRIQTVGLASYDHOSSGTFLEQAGSGIYHVKGPG 180
 QY 181 FTQOPAPNLYVSCPIYMDPKCPADPMIAFHSNDIWSINIVTREERLLTYVNE 240
 DB 181 FTQOPAPNLYVSCPIYMDPKCPADPMIAFHSNDIWSINIVTREERLLTYVNE 240
 QY 241 ANNEEDARSAGVATFVLOEEDFYSGYWCPKAEFTPSGGKILILEENDESEVEI 300
 DB 241 ANNEEDARSAGVATFVLOEEDFYSGYWCPKAEFTPSGGKILILEENDESEVEI 300
 QY 301 TSPMLETRADSPRYPTGTANPKVTFKMSIMIDAGRILIDVIXELIQFELLFEGVE 360
 DB 301 TSPMLETRADSPRYPTGTANPKVTFKMSIMIDAGRILIDVIXELIQFELLFEGVE 360
 QY 361 YIYRAGWTEGKYAMSTILDRSQRLQIVLISPLFTFVDDVWERQLIESVDSVPL 420
 DB 361 YIYRAGWTEGKYAMSTILDRSQRLQIVLISPLFTFVDDVWERQLIESVDSVPL 420
 QY 421 ILYEETDWINIHIDIFHVPFQSHHEIEFIFASECTGPRHLKYTISI 469
 DB 421 ILYEETDWINIHIDIFHVPFQSHHEIEFIFASECTGPRHLKYTISI 469

RESULT 10

Q9HBM3 PRELIMINARY; PRT; 465 AA.
 ID Q9HBM3
 AC Q9HBM3;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Dipeptidyl peptidase 8 (Fragment).
 GN DPP8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=20467194; PubMed=11012666;
 RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
 RA Gortel M.D.;
 RA "Cloning, expression and chromosomal localization of a novel human
 RT dipeptidyl peptidase (DPP) IV homology, DPP8";
 RL Eur. J. Biochem. 267:6140-6150(2000).
 DR EMBL; AF221636; AAC29768.1; -
 DR MEROPS; S09.018; -
 KW Hypothetical protein.
 FT NON_TER 1
 SO SEQUENCE 465 AA; 53197 MW; 229399EC0A4FE29CB CRC64;

Query Match 51.5%; Score 2422; DB 4; Length 465;
 Best Local Similarity 82.3%; Pred. No. 9.2e-180;
 Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 318 TGTANPKVTFKMSIMIDAGRILIDVIXELIQFELLFEGVEIYARAGWTEGKYANSI 377
 DB 1 TGTANPKVTFKMSIMIDAGRILIDVIXELIQFELLFEGVEIYARAGWTEGKYANSI 377
 QY 378 ILDRSQRLQIVLISPLFTFVDDVWERQLIESVDSVPLIYEETDWINIHIDIF 437
 DB 61 ILDRSQRLQIVLISPLFTFVDDVWERQLIESVDSVPLIYEETDWINIHIDIF 437
 QY 438 HVFPQSHHEIEFIFASECTGPRHLKYTISILKSKYKSSGGIAPASDCKPIKEIA 497
 DB 121 HVFPQSHHEIEFIFASECTGPRHLKYTISILKSKYKSSGGIAPASDCKPIKEIA 497
 QY 498 ITSGEMVLAGHNSINQVDEVRRLVYEGTQSLPHEHLVYVSVNPGVTRLDGYSH 557
 DB 181 ITSGEMVLAGHNSINQVDEVRRLVYEGTQSLPHEHLVYVSVNPGVTRLDGYSH 557
 QY 558 SCCISQCHDFISKYSNOKNPHCVSLKLSBPEDDPTCKTEFWATTIDSGPLDPYTP 617
 DB 241 SCCISQCHDFISKYSNOKNPHCVSLKLSBPEDDPTCKTEFWATTIDSGPLDPYTP 617
 QY 618 EIFSESTGFTLYGMLYKPHDLPKXKPTVLPYIGAPQVQVYVNNFEGVYRLNTLA 677
 DB 301 EIFSESTGFTLYGMLYKPHDLPKXKPTVLPYIGAPQVQVYVNNFEGVYRLNTLA 677
 QY 678 SLGVVVVVIDNRSGCHGKLFEGAFKYMQOIEIDDOVGLQYLASRDPIDLRVGHG 737
 DB 341 ----- 737
 QY 738 MSYGYLSLMAHQSDIFRVAIAGAPVTLMIFPDGYTERVYMGHPONEGYLGSYAM 797
 DB 341 -----VAIAGAPVTLMIFPDGYTERVYMGHPONEGYLGSYAM 797
 QY 798 QAEKFPSEPNRLILHGFLENVFAFSTISLSFLVAKGPRYDQIYPOEHSIRVPSG 857
 DB 381 QAEKFPSEPNRLILHGFLENVFAFSTISLSFLVAKGPRYDQIYPOEHSIRVPSG 857
 QY 858 EHYELHLHLQENLSRIALAKYI 882
 DB 441 EHYELHLHLQENLSRIALAKYI 465

RESULT 11

Q8N3F5 PRELIMINARY; PRT; 628 AA.
 ID Q8N3F5
 AC Q8N3F5;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN DKFZP762M2413.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Melanoma;
 RA Ansoyge W., Winkner U., Mewes H.W., Weil B., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL834376; CAD39039.1; -
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR000379; Ser-estrs_site.
 DR Pfam; PF00326; Peptidase_S9; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SO SEQUENCE 628 AA; 71368 MW; DC0B6A9440507C3B CRC64;

Query Match 44.0%; Score 2068; DB 4; Length 628;
 Best Local Similarity 60.6%; Pred. No. 5e-152;

Query Match 35.0%; Score 1645.5; DB 4; Length 310;
 Best Local Similarity 86.4%; Pred. No. 1.2e-119;
 Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

QY 524 FEETKSPLEHLYVSVYVNGEVTLRDRGSHSCCISQHCDFISKXSNQKPCVSL 583
 DB 1 FEGTKSPLEHLYVSVYVNGEVTLRDRGSHSCCISQHCDFISKXSNQKPCVSL 60
 QY 584 YKLSDEDDPTCKTEFWATILDSAGPLDYPPEIFSESTTGFTLYGMLYKPHLOPG 643
 DB 61 YKLSDEDDPTCKTEFWATILDSAGPLDYPPEIFSESTTGFTLYGMLYKPHLOPG 120
 QY 644 KKPFTLYFYGGQVVLNNRFGKVKYFRNLNTLASLGYYVVTIDNRGSGHGLKFGAERK 703
 DB 121 KKPFTLYFYGGQVVLNNRFGKVKYFRNLNTLASLGYYVVTIDNRGSGHGLKFGAERK 134
 QY 704 YKMGQIEIDDOVEGLQYIASRYDFIDLDVGVGHGMSYGYLSMALMORSDFRVAIAGA 763
 DB 135 ---GQIEIDDOVEGLQYIASRYDFIDLDVGVGHGMSYGYLSMALMORSDFRVAIAGA 191
 QY 764 PVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGFLENVHFA 823
 DB 192 PVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGFLENVHFA 251
 QY 824 HRSILSFLVRAKPYDLOIYQERHSIRVPSGGEHYELHLHYLGENTGSRIALKVI 882
 DB 252 HRSILSFLVRAKPYDLOIYQERHSIRVPSGGEHYELHLHYLGENTGSRIALKVI 310

RESULT 14

075273 PRELIMINARY; PRT; 508 AA.

ID 075273; PRELIMINARY; PRT; 508 AA.
 AC 075273;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE R26984.1 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
 RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
 RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
 RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
 RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
 RA Krommiller B., Arellano A., Montgomery W., Ow D., Nolan W., Trong S.,
 RA Kobayashi A., Olsen A.S., Carrano A.V.;
 RT "Sequence analysis of a 2.5 Mb region in 19p13.3";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC005594; AAC33801.1; -
 DR MEROPS; S09.019; -
 DR InterPro; IPR000379; Set_estrs_site.
 FT NON_TER 1
 SQ SEQUENCE 508 AA; 57750 MW; 2FAD6645BE2D2C89 CRC64;

Query Match 35.0%; Score 1643.5; DB 4; Length 508;
 Best Local Similarity 57.6%; Pred. No. 3.8e-119;
 Matches 313; Conservative 77; Mismatches 108; Indels 45; Gaps 6;

QY 340 TIDVIDKELIQPFELIEGVEYIARAGWTPEBGKVASILDRSQTRLOIVLISPELTPV 399
 DB 1 IVSTQEKELVQPFSSLPKVEYIARAGWTPEBGKVASILDRSQTRLOIVLISPELTPV 53
 QY 400 EDDVNERQLLESVDSTPLIIVBETDININIDIRHVPOSH-EEIEIFIFASECKT 458
 DB 54 TENEBQRIASAPAVPRNVQPVVVEVTVNINVDIYPPQSGEDELFLRANCKT 113
 QY 459 GFRHLKYITSLKSKYKSSGGLPAPSDFCPIKEEIAITSGEVVLGRHSNIGVDEV 518

DB 114 GFCHLYKVAIVLKSGQYDWSPPSPG-----EGEQSLTNAIW-----VNEE 154
 QY 519 RRLVYEGTKDPSLEHLYVSVYVNGEVTLRDRGSHSCCISQHCDFISKXSNQKPCVSL 578
 DB 155 TKLVYFGTKDPSLEHLYVSVYVNGEVTLRDRGSHSCCISQHCDFISKXSNQKPCVSL 214
 QY 579 HCVSLYKLSDEDDPTCKTEFWATILDSAGPLDYPPEIFSESTTGFTLYGMLYKPH 638
 DB 215 PCHVYKLSDEDDPTCKTEFWATILDSAGPLDYPPEIFSESTTGFTLYGMLYKPH 265
 QY 639 DLQPGKKTFTLYFYGGQVVLNNRFGKVKYFRNLNTLASLGYYVVTIDNRGSGHGLK 698
 DB 266 ALQPGKKTFTLYFYGGQVVLNNRFGKVKYFRNLNTLASLGYYVVTIDNRGSGHGLK 325
 QY 699 BGAFTKMGQIEIDDOVEGLQYIASRYDFIDLDVGVGHGMSYGYLSMALMORSDFRVAIAGA 758
 DB 326 BGAFTKMGQIEIDDOVEGLQYIASRYDFIDLDVGVGHGMSYGYLSMALMORSDFRVAIAGA 385
 QY 759 ALAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGFLENVHFA 818
 DB 386 ALAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGFLENVHFA 445
 QY 819 NVHFAHTSLSFLVRAKPYDLOIYQERHSIRVPSGGEHYELHLHYLGENTGSRIALKVI 869
 DB 446 NVHFAHTSLSFLVRAKPYDLOIYQERHSIRVPSGGEHYELHLHYLGENTGSRIALKVI 505
 QY 870 ENL 872
 DB 506 EYL 508

RESULT 15

09BVR3 PRELIMINARY; PRT; 439 AA.

ID 09BVR3; PRELIMINARY; PRT; 439 AA.
 AC 09BVR3;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strusberg R.;
 RA Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC000970; AA000970.1; -
 DR MEROPS; S09.019; -
 DR Gene; HGNC:18648; DPP9.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR000379; Set_estrs_site.
 DR Pfam; PF00326; Peptidase_S9; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 439 AA; 49926 MW; A18BBA9E12092BAF CRC64;

Query Match 34.3%; Score 1610.5; DB 4; Length 439;
 Best Local Similarity 67.0%; Pred. No. 1.1e-116;
 Matches 294; Conservative 57; Mismatches 87; Indels 1; Gaps 1;

QY 435 DIFHVFPOSH-EEIEIFIFASECKTGFRHLKYITSLKSKYKSSGGLPAPSDFCPIK 493
 DB 1 DIFHVFPOSH-EEIEIFIFASECKTGFRHLKYITSLKSKYKSSGGLPAPSDFCPIK 60
 QY 494 EEIATSGEWEVVLGRHSNIGVDEVRLVYEGTKDPSLEHLYVSVYVNGEVTLRDR 553
 DB 61 EEIATSGEWEVVLGRHSNIGVDEVRLVYEGTKDPSLEHLYVSVYVNGEVTLRDR 120
 QY 554 GYSHSCCISQHCDFISKXSNQKPCVSLYKLSDEDDPTCKTEFWATILDSAGPLD 613

```

Db      121 GFSHSCSMSONFDMFVSHYSSVSTPPCVHYKLSGPDDDLHKOPRFWASMEASCPD 180
QY      614 YPEPEIJSFESTTGFITLGMLYKPHDLOPKKXPTVLYIGSPQVULVNNRFKGVKTEL 673
Db      181 YPEPEIHFHFRSDVRLYGMLYKPHALQPKKHPTVLYVGGPOVULVNNSPFKGIXKXRL 240
QY      674 NTLASLGVVVVVIDNRSCHRGKLFEGAFYKMGQLEIDDOVEGLQYLAARYDFIDLDY 733
Db      241 NTLASLGVAVVVIDGRSCCORGLRFEGALKNQMGQVEIEDQVEGLQFAEKYGFIDLSRV 300
QY      734 GINGMSYGYLSLMAIMORSIDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNGYYLG 793
Db      301 AINGMSYGGFLSLMGLIHKPQVFKVAIAGAPVTVMAYDTGYTERYMDVPENNQHGYEAG 360
QY      794 SVAMQABKPPSEPNRLLILHGFIDENVHFAHTSILSFLVRAGKPYDIQIYQERHSIRY 853
Db      361 SVALHYEKLPENEPNRLLILHGFIDENVHFAHTSILSFLVRAGKPYDIQIYQERHSIRC 420
QY      854 PESGEHYEHLHLHYLOENTL 872
Db      421 PESGEHYEVTLLHFLQBYL 439

```

Search completed: October 15, 2003, 17:13:41
 Job time : 61.4095 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:15:46 ; Search time 9629.41 seconds

(without alignments)
3747.092 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700
Sequence: 1 MAAAMETGQGVETFTADG.....HLHYLOENIGSRITALKVI 882

Scoring table:
BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Rgapop 6.0 , Rgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1h
-O=/cgn2.1/USPTO.spool/US10070464/runat.15102003.113553.24810/app.query.fasta_1.2652
-DB=Genembl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCUT=0 -LOOPEXT=0
-UNITS-bits -START=1 END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464@cgn.1.13561@runat.15102003.113553.24810 -NCPU=6 -ICPU=3
-NO MMAP -LARGESUBSTRY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -RGAPOP=6
-RGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Genembl:*
1: gb Da:*
2: gb Hcg:*
3: gb In:*
4: gb Om:*
5: gb Ov:*
6: gb Pat:*
7: gb Ph:*
8: gb Pl:*
9: gb Pr:*
10: gb Ro:*
11: gb Sts:*
12: gb Sy:*
13: gb Un:*
14: gb Vi:*
15: em Ba:*
16: em Fun:*
17: em Hum:*
18: em In:*
19: em Mu:*
20: em Om:*
21: em Or:*
22: em Ov:*
23: em Pac:*
24: em Ph:*
25: em Pl:*
26: em Ro:*
27: em Sts:*
28: em Un:*

29: em Vi:*
30: em Htg Hum:*
31: em Htg Inv:*
32: em Htg Other:*
33: em Htg Mus:*
34: em Htg Pln:*
35: em Htg Rod:*
36: em Htg Mam:*
37: em Htg Vrt:*
38: em Sy:*
39: em Htgo Hum:*
40: em Htgo Mus:*
41: em Htgo Other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4700	100.0	2649	6 AX354795	AX354795 Sequence
2	4700	100.0	2656	9 AY172659	AY172659 Homo sapi
3	4700	100.0	2671	6 AX608725	AX608725 Sequence
4	4700	100.0	3106	6 AX342633	AX342633 Sequence
5	4700	100.0	3127	9 AF221634	AF221634 Homo sapi
6	4700	100.0	3143	6 AX354793	AX354793 Sequence
7	4696	99.9	4535	9 BC040203	BC040203 Homo sapi
8	4680	99.6	4829	6 AX608735	AX608735 Sequence
9	4536.5	96.5	4799	10 BC043124	BC043124 Mus muscu
10	4397.5	93.6	3130	6 BC030688	BC030688 Homo sapi
11	4385.5	93.3	4685	6 AX608745	AX608745 Sequence
12	4385	93.3	4676	6 AX608743	AX608743 Sequence
13	4118	87.6	2842	6 AX405770	AX405770 Sequence
14	4092.5	87.1	4523	6 AX608731	AX608731 Sequence
15	4091	87.0	2778	6 AK000290	AK000290 Homo sapi
16	3970.5	84.5	2510	6 AX338497	AX338497 Sequence
17	3771	80.2	2668	6 AX405771	AX405771 Sequence
18	3661.5	77.9	4309	6 AX608737	AX608737 Sequence
19	3364.5	71.6	2161	6 BD157001	BD157001 Primer fo
20	3364.5	71.6	2161	9 AK027826	AK027826 Homo sapi
21	2870	61.1	2602	9 AY172660	AY172660 Homo sapi
22	2870	61.1	2617	6 AX608727	AX608727 Sequence
23	2870	61.1	3716	6 AX480934	AX480934 Sequence
24	2870	61.1	4219	6 AX608751	AX608751 Sequence
25	2870	61.1	4232	9 BC037948	BC037948 Homo sapi
26	2870	61.1	4295	9 AF452102	AF452102 Homo sapi
27	2870	61.1	4302	6 AX608747	AX608747 Sequence
28	2867	61.0	2676	6 AX524928	AX524928 Sequence
29	2863	60.9	3000	9 AF542510	AF542510 Homo sapi
30	2820.5	60.0	4180	6 AX608759	AX608759 Sequence
31	2820.5	60.0	4263	6 AX608757	AX608757 Sequence
32	2649	56.4	4076	6 AX608755	AX608755 Sequence
33	2649	56.4	4159	6 AX608753	AX608753 Sequence
34	2599.5	55.3	4037	6 AX608763	AX608763 Sequence
35	2599.5	55.3	4120	6 AX608761	AX608761 Sequence
36	2422	51.5	1669	9 AF221636	AF221636 Homo sapi
37	2139.5	45.5	3243	9 HSM80548	AL834576 Homo sapi
38	1836.5	39.1	1083	9 AF221637	AF221637 Homo sapi
39	1730.5	36.8	2261	6 AX713363	AX713363 Sequence
40	1730.5	36.8	2261	9 AK054656	AK054656 Homo sapi
41	1645.5	35.0	1197	9 AF221635	AF221635 Homo sapi
42	1610.5	34.3	2546	9 BC000970	BC000970 Homo sapi
43	1596	34.0	3740	3 BT001499	BT001499 Drosophi1
44	1391	29.6	2411	6 AX608749	AX608749 Sequence
45	1360	28.9	215711	3 AC008205	AC008205 Drosophi1

RESULT 1

ALIGNMENTS

AX354795
 LOCUS AX354795 2649 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 3 from Patent WO0179473.
 ACCESSION AX354795
 VERSION AX354795.1 GI:18619528
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 Meyers, R.A. and Williamson, M.
 21953, a human prolyl oligopeptidase family member and uses thereof
 Patent: WO 0179473-A 3 25-OCT-2001;
 Millennium Pharmaceuticals, Inc. (US)
 FEATURES
 source 1..2649
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 803 a 514 c 586 g 746 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 2649
 Score: 4700.00 Matches: 882
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0
 DB: 6
 US-10-070-464-1 (1-882) x AX354795 (1-2649)

QY 1 MetAlaAlaIaMeGluThrGluGluLeuGlyValGluIlePheGluThrAlaAspCys 20
 DB 1 ATGGAGAGAGCAATGGAAACAGACAGCTGGGTGTGATATTGAAACCTCCGACGCT 60
 QY 21 GluGluAsnIleGluSerGlnAspArgProlysluGluProPheTyrValGluArgTyr 40
 DB 61 GAGGAGATATTGATCAACAGATCGGCTTAATTGGAGCCCTTTTATGTTGAGGCGGTAT 120
 QY 41 SerTPSerGlnLeuValysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
 DB 121 TCCTGGAGTACGCTTAAAGCTGCTTCCCATACCAAAATATCATGCTACATGATG 180
 QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
 DB 181 GCTAAGGACCAACATGATTCATGTTTGTGAAGAGATGATCCAGATGACCTCATTTCA 240
 QY 81 AspArgLysTyrTyrLeuAlaMetSerGlyLysAsnArgLysAsnThrLeuPheTyrSer 100
 DB 241 GACGAATCTATTAACCTTGCATGCTGTGTGAAGAGATGATCCAGATGACCTCATTTCT 300
 QY 101 GluIleProlyslThrIleAsnArgAlaAlaValLeuMetLeuSerTyrProlyslProleu 120
 DB 301 GAATTCCTCCAAACATTAATTAAGACAGACATCTTAAGCTCTCTTGAAGCCCTCTTTG 360
 QY 121 AspleuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgLysGluGluLeuLeuArg 140
 DB 361 GATCTTTTTCAGGACACCTGACCTATGGAATGTATTCGAGAAAGAAAGAACTTTAAGA 420
 QY 141 GluArgLysArgIleGlyThrValGlyIleAspTyrAspTyrHisGlnGlySerGly 160
 DB 421 GAAAGAAACGCAATTTGAAACGCTGGAATTCCTCTTTCATTCACCAAGAGAGTGA 480
 QY 161 ThrPheLeuPheGlnAlaGlySerGlyTyrHisValLysAspGlyGlyProGlnGly 180
 DB 481 ACATTTCTGTTTCAAGCCGGTAGTGAAATTATATACGTAAGAAAGTGGGCGCAAGAGA 540
 QY 181 PheThrGlnGlnProleuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
 DB 541 TTAAAGCAACACTTTAAGGCCCAATCTAGTGAAACTGATTGTCTCCCAATACGATG 600

QY 201 AspProlysluGluValAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
 DB 601 GATCAAAATATATGCTCCGCTGATCCAGACCTGATGCTTTTATACATAGCAACATATT 660
 QY 221 TyrIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
 DB 661 TGGATATCTAACATCGTAAACAGAGAAAGAGAGACTCATTAATGTGCAATATGAGCTA 720
 QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGlu 260
 DB 721 GCCAACATGGAAGAGAGATGACAGATCGCTGAGTGGCTGCTTGTCTCCAAAGAGAA 780
 QY 261 PheAspArgTyrSerGlyTyrTyrTyrProCysProLysAlaGluThrProSerGlyGly 280
 DB 781 TTGATATGATATTTCTGCGTATTTGGTGGTGTCCAAAGCTGAAACACTCCAGTGTGTGT 840
 QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
 DB 841 AAAATTTCTTGAATTCATATGAAGAAATATGATGATGTGAGTGAATTAATTCATGTT 900
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
 DB 901 ACATCCCTATGTTGGAACAGAGAGGCGCAATTCATTCGTTATCTTAAACAGGTACA 960
 QY 321 AlaAsnProlyslValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
 DB 961 GCAATCTTAAGACACTTTTAAGATGTCAGAAATTAATGATGATGCTGAAGAGAGATC 1020
 QY 341 IleAspValIleAspLysGluLeuIleGluProPheGluIleLeuPheGluGlyValGlu 360
 DB 1021 ATAAATGTCATGATGAAGAGAACTAATTCACCTTTTGATTCATTTGAAGAGGTGAA 1080
 QY 361 TyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIleLeuLeuAsp 380
 DB 1081 TATATTGCGACAGCTGATGATGATCTCTGAGGAAATATGCTTGGTCCATCTCAATAT 1140
 QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
 DB 1141 CGCTCCAGACTCGCTGACAGATGATGTTGATTCACCTGATATTATTATCCAGATAGA 1200
 QY 401 AspAspValMetGluArgGluArgLeuIleGluSerValProAspSerValThrProleu 420
 DB 1201 GATGATGTTATGAGAAAGCAGAGACTAATGATGATGAGTCTGATTCGTATCCGACCTA 1260
 QY 421 IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
 DB 1261 ATTATCTATGAGAAACAGACAGATCTGATTAATATCCATGACATTTTCACTTTT 1320
 QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
 DB 1321 CCCCAAGTCCAGAAAGAGAAATGAGTTTATTTTGGCTGATGCAAAACAGGTTTC 1380
 QY 461 ArgHisLeuTyrLysIleThrSerIleLeuIleGluSerLysTyrLysArgSerSerGly 480
 DB 1381 CGTATTTATACAAATTAATCACTTATTTTAAAGAAAGCAATTAACGATCCAGTGT 1440
 QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
 DB 1441 GGGCTGCTGCTCCAAATGATTCATAGTGTCTATCAAGAGAGAGATGACATTAACAGT 1500
 QY 501 GlyIleTyrProIleValLeuGlyValArgHisGlySerAsnIleGlnValAspGluValArgArg 520
 DB 1501 GGTGAATGGAGAGGTTCTTGGCCGCGATGATCTAATATCCAAAGTTGATGAAGTACAGAG 1560
 QY 521 LeuValTyrPheGluGlyThrLysAspSerProleuGlnHisHisLeuTyrValValSer 540
 DB 1561 CTGGTATATTTTGAAGCACAACAAAGATCCCTCTTAAGCATACCTGTACGTATGATG 1620
 QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgLysTyrSerHisSerCysCys 560
 DB 1621 TACGTAATATCTGAGAGAGTGAACAAGGCTGATCAGCTGGCTATCAATTTCTTGTCTGC 1680
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580

```

Db      1681 ATCAGTACGACCTGCTGCTTTTATATAGTATGATGATACCAAGAAATCCACACTGT 1740
Qy      581 ValSerLeuTyrlsLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db      1741 GTGTCCTTTACAGAGCATATCAAGCTCGTGAAGATGACCCAACTTGCAGAAACAAAGAAATT 1800
Qy      601 TTPAlaThrLleLeuAspSerAlaGlyProLeuProAspTyrlThrProProGluLlePhe 620
Db      1801 TGGCCCAACCAATTTGGATTTCAGCAGGCTCTCTCTGACTATCTCCCTCCAGAAATTTTC 1860
Qy      621 SerPheGluSerThrThrGluPheThrLeuTyrlGlyMetLeuTyrlLysProHisAspLeu 640
Db      1861 TCTTTGAAGACTACTGAGATTGATTCATTGATGAGAGCTCTCAAGAGCTCATGATCTTA 1920
Qy      641 GluProGlyLysLysTyrlProThrValLeuPheLleTyrlGlyGlyProGluValGluLeu 660
Db      1921 CAGCTCGAAAGAAATATCTACTGCTGCTGTTATATATGTTGATCTCCAGGTCAGATGTG 1980
Qy      661 ValAsnAsnArgPheLysGlyValLysTyrlPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db      1981 GTGAAATATCGGATTAAAGAGGTCAGATATTTCCGCTTGATATACCTTAGCTCTAGGT 2040
Qy      681 TyrlValValValLleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGly 700
Db      2041 TATGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
Qy      701 AlaPheLysTyrlLysMetGlyGluLleGluLleAspAspGluValGluGlyLeuGluLlyr 720
Db      2101 GCCTTTAAATATATAATGCGGTCAATATGAAATGACATGCGTGAAGAGCTCCAAATAT 2160
Qy      721 LeuAlaSerArgTyrlAspPheLleAspLeuAspArgValGlyLleHisGlyTyrSerTyrl 740
Db      2161 CTAGCTCTTCGATATGATATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 2220
Qy      741 GlyLysLysLeuSerLleLeuMetAlaLeuMetGluArgSerAspLlePheArgValAlaLle 760
Db      2221 GAGAGATACCTCTCTCGATGATGATTAATGACAGGCTCAGATATCTTCAGGATGCTAAT 2280
Qy      761 AlaGlyAlaProValThrLeuTyrlPheTyrlAspThrGlyTyrlThrGluArgTyrlMet 780
Db      2281 GCTGGGGCCCCAGTCACTCTGTGATCTTCTATGATACAGATACAGGAAACGTTATATG 2340
Qy      781 GlyHisArgProGluAsnGluGluGlyTyrlTyrlLeuGlySerValAlaMetGluAlaGlu 800
Db      2341 GGTCACCTCAGACAGATGAAACAGGCTATTACTAGATCTGTGGCCATGCAAGCAGAA 2400
Qy      801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db      2401 AAGTTCCTCTGAAACCAATCGTTTCTGCTTCAATGCTTCTGAGATGAGAAATGTC 2460
Qy      821 HisPheAlaHisThrSerLleLeuLeuSerPheLeuValArgAlaGlyLysProTyrlAsp 840
Db      2461 CATTTTCACATACCACTATATATTAATGATTTTATGAGGGCTGGAAGACCATATGAT 2520
Qy      841 LeuGluLleTyrlProGluGluArgHisSerLleArgValProGluSerGlyLysHisTyrl 860
Db      2521 TTTACAGATCTATCTCAGAGAGACACAGATTAAGATTCCTGATGAGGAGAACATTAT 2580
Qy      861 GluLeuHisLeuLeuHisTyrlLeuGluGluAsnLeuGlySerArgLleAlaLleLeuLys 880
Db      2581 GAACCTGATCTTTTGGACTACCTTCAAGAAACCTTGATCAAGATTCGATTCGCTTAAA 2640
Qy      881 ValLle 882
Db      2641 GTGATA 2646

```

```

RESULT 2
LOCUS   AY172659      2656 bp      mRNA      linear      PRI 08-JAN-2003
DEFINITION Homo sapiens dipeptidyl peptidase IV-related protein-1 (DPP1)
          mRNA, complete cds.
ACCESSION AY172659

```

```

VERSION AY172659.1 GI:27549549
KEYWORDS
SOURCE  Homo sapiens (human)
ORGANISM
REFERENCE
  1 (bases 1 to 2656)
  Mammalia; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Euteleostomi; Primates; Catarrhini; Homnidae; Homo.
AUTHORS  Qi,S., Akinaanya,K., Riviere,P. and Junien,J.-L.
TITLE     Novel Serine Protease Genes Related to Dppiv
JOURNAL  Patent: US (WO 02/1134)-A 18-APR-2002;
          2 (bases 1 to 2656)
AUTHORS  Qi,S., Akinaanya,K., Riviere,P. and Junien,J.-L.
TITLE     Direct Submission
JOURNAL  Submitted (04-NOV-2002) Ferring Research Institute, 3550 General
          Atomics Ct., San Diego, CA 92121, USA
FEATURES
  source
    1..2656
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /chromosome="15"
      /map="15q22.1-q22.2"
      1..2656
      /gene="DPP1"
      /gene="DPP1"
      8..2656
      /gene="DPP1"
      /product="dipeptidyl peptidase IV-related protein-1"
      /protein_id="AA017261.1"
      /db_xref="GI:27549549"
      /translation="MAAMETEQLEVEFTADCEENIESQDRPKLEPFVERYSWG
      LKULADTRRYHYMAKADHDFVNRNDPDGSHRIYILAMSGENRENTLFSYSEI
      PKTINRAVALMLSWKPLDLDFQATLDVMDGSEELREKRIGTVAISADYQSGG
      TLFQSGGTYHYVGDGPGFOTPOPLPNI.VERTSCPIRMNDPCLPDPDWIATFHSN
      DWISNIVTREERERLTYYHNEILANMEDASAGATVRLDREPRVSGYGCPCRAET
      PSQKILRIIYENDESEVATHTSMTLTRADSRYPRTGTANPKVTRKSEIMI
      DABGRIITVDIKELIPEFELFEGVEYIAAGWTPREKYAWSILLDSQRLQVLIS
      PELFIPEVDVMEROLIESVPDSVTEPLIYEETTDRIWHDFHVFQSHSEIER
      IFASECTQGRHLVYKITSILKESKYKSSGGLPSPDFKCPIKEEIAITSEMEVLR
      HGSNIQVDEVRRLVYEGTKDSPLEHLVYVSVYNPEVTRLDRGVSHSCISQHC
      FPKSKYNOCKPHCVSLYKLSPEDDPTCTKTKEMATILDSAGLPRYTPETISFES
      TTGPTLGMILKPHDLOPKKTYPLVLTYGSPQVQLVNNRFGKVFRLNLASGLV
      VVVIDNRGSHRGKKEGAFKRYKQGLIIDDQVGLQYLASRYPIDLRGHSWMSY
      GGYLSMALMQRSDIFRVALAGAPVTIMLIFYDGYTRMYGHPDNRGGLVLSVAMQ
      AEKFPSEPNRLLLHGLDENVHFHAFHISILSLFLVRAGKPYDQIYPOERSIRVPS
      GEHYELHLHYEINDGSRITAAKVT"
BASE COUNT  804 a 517 c 588 g 747 t
ORIGIN
Alignment Scores:
  Pred. No.: 0
  Score: 4700.00
  Percent Similarity: 100.00%
  Best Local Similarity: 100.00%
  Query Match: 100.00%
  DB: 9
  Gaps: 0
US-10-070-464-1 (1-882) x AY172659 (1-2656)
Qy      1 MetAlaAlaAlaMetGluThrGluGluGluValGluLlePheGluThrAlaAspCys 20
Db      8 ATGCGACGACCAATGGAACAGACAGCTGGGTGAGATATTTGAAACCTGCGGAGATGT 67
Qy      21 GluGluAsnLleGluSerGluAspArgProLysLeuGluThrPheTyrlValGluArgTyrl 40
Db      68 GAGAGAAATTTTAATACACGAGATCGCTTAATTTGAGGCTTTTATGTTGAGCGGATAT 127
Qy      41 SerTrpSerGluLeuLysLysLeuLeuAlaAspThrArgLysTyrlHisGlyTyrlMetMet 60
Db      128 TCCTGAGGTACGCTTAAAGAGCTGCTGCGGATACCAAGAAATATCATGCTACATGATG 187
Qy      61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80

```


QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db 2408 AAGTTCCCTCTGTAACCAAAACGTTTACTGCTTACATGGTTTCTCGAAGCAATGTC 2467
QY 821 HisPheLeuHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
Db 2468 CATTTCGACATACCACTATATTTACTAGTTTTTTATGAGGCGCTGAAACCCATAGAT 2527
QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
Db 2528 TTACAGATCTATCTCTCAGAGAGACACAGCATAGAGTCTCTGAATCGGAGAACATTAT 2587
QY 861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaLeuLys 880
Db 2588 GAATCGATCTTTTTCGACTACTACTTCAAGAAACCTTGATCATCGTATGCTGCTAAAA 2647
QY 881 ValIle 882
Db 2648 GTGATA 2653

RESULT 3
AX608725 2671 bp DNA linear PAT 17-FEB-2003
LOCUS Sequence 2 from Patent WO0231134.
DEFINITION AX608725
VERSION AX608725.1 GI:28404298
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
AUTHORS Novel serine protease genes related to dprlv
TITLE Patent: WO 0231134-A 2 18-Apr-2002;
JOURNAL Perring BV (NL)

FEATURES
source Location/Qualifiers
1..2671
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 805 a 524 c 594 g 748 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 2671
Score: 4700.00 Matches: 882
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 6

US-10-070-464-1 (1-882) x AX608725 (1-2671)

QY 1 MetValAlaIleMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
Db 8 ATGGCACAGCAATGAGAAACAGAACACCTGGGTGTAGATATTGAAACCTGCGAGCTGT 67
QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValAlaArgTyr 40
Db 68 GAGAGAAATTTGATTCACAGGATCGGCTTAAATTGAGCCTTTTATGTTGAGCGGAT 127
QY 41 SerTyrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db 128 TCCGAGGTCAAGCTTTAAAGCTGCTCCGATACCGAAATATCATGCTCATCATATG 187
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db 188 GCTAAGCACACATGATTTTCATGTTTGAGAGGAAATGATCCAGATGAGCCTCATTTA 247
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
Db 248 GACGATCTATATCTTCCATGCTGTGTGAGAAACAGAAATATACCTGTTTATCTT 307

QY 101 GluIleProLysThrIleAsnArgAlaIleValLeuMetLeuSerTyrPlyProLeuLeu 120
Db 308 GAAATTCACCAAACTATCATATAGAGCAGAGCTTTATAGCTCTCTCGAAGCCCTTTTG 367
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
Db 368 GATCTTTTTCAGGCAACACCTGACTATGATATGATCTCGAAGAAACATTTATAGA 427
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
Db 428 GAAGAAAGACATTGAAACAGTCGGAATTCCTTTACGATTTACCAAGAGAGTGA 487
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
Db 488 ACATTTCTGTTTCAAGCCGGATGATTTATACCGTTAAAGATGAGGGCCCAAGGA 547
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db 548 TTTCAGCAACACCTTTAAGCCCAATCTAGTGAAACTAGTTTCCCAACATACGATG 607
QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
Db 608 GATCCAAAATTTATGCCCTGCTGATCCAGACTGGAATGCTTTATACATAGCAAGATTT 667
QY 221 TrrIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
Db 668 TGGATATCTAATCATCTGATACCGAGAAAGAGACTCACTTATGTGCAACATGACCTA 727
QY 241 AlaAsnMetGluGluAsnAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
Db 728 GCCAACATGAGAAAGATGCGACATGACGTGAGATCCCTACCTTGTTCCAAGAAAGA 787
QY 261 PheAspArgTyrSerGlyTyrTrrCysProLysAlaGluThrThrProSerGlyGly 280
Db 788 TTTGATGATATTTCTGCTATTTGATGAGAAATGATGATCTGAGGTAATTTATCATGTT 847
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleHisVal 300
Db 848 AAAATTCCTTAAGATTTCTATATGAAAGAAATGATGATCTGAGGTAATTTATCATGTT 907
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db 908 ACATCCCTTATGTTGAGAAACAGAGGCGAGATTCCTTATCTTAAACAGGTACA 967
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db 968 GCAATATCTTAAGCTACTTTTAAAGATGTCAGAAATTAATGATGCTGTAAGAAAGATC 1027
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db 1028 ATGATGTCATATGATTAAGAACTAATCAACCTTTTGAGATTCTAATTGAAAGAGTTGAA 1087
QY 361 TyrIleAlaArgAlaGlyTrrThrProGluGlyLysTyrAlaTrrPserIleLeuLeuAsp 380
Db 1088 TATATTTCCAGAGCTGATGATGATCTCGAAGGAAATATGCTTGCTCATCTCATGAT 1147
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db 1148 CGCTCCAGACTGCTCGCAGATGATGTTGATTCCTCACTGAATTAATTTATCCAGTAA 1207
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db 1208 GATGATGTTATGAAAGGACAGACATCATGAGTCACTGCTGATTCGTGAGCCCACTA 1267
QY 421 IleIleTyrGluGluThrThrAspIleTrrIleAsnIleHisAspIlePheHisValPhe 440
Db 1268 ATTATCTATAGAAAGAAACAGACATCTGATTAATATTCATGACATCTTTCATGTTT 1327
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
Db 1328 CCCCAAGTACAGAAAGGAAATGAGTTTATTTTTCCTCTGAAATGCAAAACAGGTTTC 1387

QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
 DB 1368 CCGATTTATACAAATATACATCTATTTTAAAGAAACAAATATTAACGATCCAGTGT 1447
 QY 461 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
 DB 1448 GGGCGCTGCTCCCAAGATTTCAAGTGTCTTATCAAGAGAGATGCAATTAACGAT 1507
 QY 501 GlyLutTyrGluValLeuGlyLysHisGlySerAsnIleGlnValAspGlyValArgArg 520
 DB 1508 GGTGAATGGAAAGTCTTGCGCGGCAATGATTAATATCCAAAGTTGATGAAGTCAAGG 1567
 QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValLysSer 540
 DB 1568 CTGGTATATTTGAAGGACCAACAAAGATCCCTTTAGAGCATCACCTGTACGTACTAGT 1627
 QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
 DB 1628 TACGTAATCCGAGAGGTGACAAAGGCTGACTGACCGTGGCTACTCACTTCTTGCTGC 1687
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
 DB 1688 ATCAGTCGACACTGACTGCTTTTATAGTAAGTATAGTACCAAGAAAGATCCACACTGT 1747
 QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
 DB 1748 GTGTCCCTTTACAACTATCAAGTCTCTAGAGTGAACCAACTGCAAAACAAAGCAATTT 1807
 QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspGlyTyrThrProGluIlePhe 620
 DB 1808 TGGGCGACATTTGGATTGACAGAGGTCTCTCTGACATATACCTCCCGAATTTTC 1867
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
 DB 1868 TCTTTTGAAGACTACTGATGATTTACATTTGATGGATGCTCTACAGCCCTCATCATCTA 1927
 QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
 DB 1928 CAGCCTGGAAGAAATATCTCACTGTGCTCATATATGGTGGTCCAGGTGCACTGTG 1987
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
 DB 1988 GTGATATATCGATTTAAAGAGTCAAGATTTCCCTTGATTAACCTTGACCTCTAGGT 2047
 QY 681 TyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
 DB 2048 TATGTGTTGTATGATAGACAAACAGGGATCTCTGACCCAGGCTTAAATTTGAAGC 2107
 QY 701 AlaPheLysTyrLysMetGlyGlnIleGluLeuAspAsnValGlnGlyLeuGlnTyr 720
 DB 2108 GCTTTAAATATTAATGGGTCAATATGAAATTTGACATCGATGGAAGGACTCCAAATAT 2167
 QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyLysGlyTyrSerTyr 740
 DB 2168 CTAGCTTCTCGATATGATTTATGATTTAGCTTAGATCGTGGGATCCAGCGCTGCTCAT 2227
 QY 741 GlyLysTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
 DB 2228 GAGAGATACCTCTCCCTGATGGCATTAATGACAGAGTCAAGATATCTTCAAGGTTCTCAT 2287
 QY 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
 DB 2288 GCTGGGGGCCAGTACTCTGTGGATCTTCTATGATACAGATACAGGAAAGTATATG 2347
 QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLysGlySerValAlaMetGlnAlaGlu 800
 DB 2348 GGTACCTCTGACCAAGATGAACAGGGCTATTACTTAGATCTGTGGCCATGCAACAGAA 2407
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
 DB 2408 AAGTTCCTCTGAAACCAATCGTTACTGCTTACATGATGTTCTCGATAGAGATATCT 2467
 QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840

DB 2468 CATTTGACATACACAGATATATCTAGATTTTATGAGAGGCTGGAAACCATATAT 2527
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
 DB 2528 TTACAGATCTATCTCTCGAGAGACACAGCATTAAGATCTGTGAATCGGGAACATAT 2587
 QY 861 GluLeuHisIleLeuHisTyrLysGlnGluAsnLeuGlySerArgIleAlaIleLys 880
 DB 2588 GACTGCATCTTTTGCACCTACCTTCAAGAAACCTTGGATCAGATATGCTGCTTAA 2647
 QY 881 ValIle 882
 DB 2648 GTGATA 2653
 RESULT 4
 AX342633
 LOCUS AX342633 3106 bp DNA linear PAT 12-JAN-2002
 DEFINITION Sequence 30 from Patent WO0198468.
 ACCESSION AX342633
 VERSION AX342633.1 GI:18152030
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 Yue, H., Elliott, V.S., Gandhi, A.R., Lal, P., Au-Young, D.,
 Tribouley, C.M., Deleogene, A.M., Baughn, M.R., Nguyen, D.B., Lee, E.A.,
 Hafealia, A., Khan, F.A., Wajia, N.K., Yao, M.G., Ju, D.A., Peterson, C.,
 Tang, Y.T., Walsh, R.T., Azimzal, Y., Ramkumar, J., Xu, Y. and Reddy, R.
 Patent: WO 0198468-A 30 27-Dec-2001;
 JOURNAL
 fncycle Genomics, Inc. (US)
 FEATURES
 source
 1. 3106
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /note="fncycle ID No: 7160544CB1"
 BASE COUNT 928 a 633 c 704 g 841 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 3106
 Score: 4700.00 Matches: 882
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-070-464-1 (1-882) x AX342633 (1-3106)
 QY 1 MetAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
 DB 203 ATGGCAGACGACAAATGAGAAACAGCTGGGTGTGGATTTTGAATCCGGACACTGT 262
 QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
 DB 263 GAGAGAAATATTTGATCAACAGATCGGCTTAATTTGAGCTTTTATGTTAGAGGGAT 322
 QY 41 SerTyrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
 DB 323 TCTGGAGTCCGCTTAAAGCTCTGTCATACCAAGAAATATCATGAGCTTACATGATG 382
 QY 61 AlalysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
 DB 383 GCTAAGGACCAACATGATTTATATGTTGTGAAGAGATGTCCAGATGACCTCATTTCA 442
 QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
 DB 443 GACAGAAATCTATACCTTGCATGCTGTGTGAGAAACAGAAATATACATGCTTTTATCT 502
 QY 101 GluIleProLysThrIleAsnArgAlaValLeuMetLeuSerTyrLysProLeuLeu 120


```

Db      503 GAAATTCCTCCAAACTATCAATAGACAGCACTCTTAATGCTCTTGTGAGAGCCCTTTTG 562
Qy      121 AspleubheglnalathrleuaspyrGlyMetTyrsesargluglugluLeuAarg 140
Db      563 GATCTTTTTCAGCGAACACTGGAAGCTATGGAATGATCTCGAGAGAGAGAACTTTAAGA 622
Qy      141 GluArglysarGllleglyThrValGlyTleAlaserTyrsAspTyrsGlnGlySerGly 160
Db      623 GAAAGAAAGCGCATGGACAGCTGGAGATGCTTCTTACGATTCATCCAAAGAGATGGA 682
Qy      161 ThrPheubheglnalaglyserGlyTleTyrsHisVallylsaspglyGlyProGlnGly 180
Db      683 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCAGCTTAAAGATGAGAGGCGCCCAAGGA 742
Qy      181 PheThrGlnGlnProleuAargProAsnLeuValGlnThrSerCysProAsnIleArgMet 200
Db      743 TTTTACGACACACTTTTAAAGGCCCAATCTAGTGGAAACTAGTTGTCCTCCACATRGGATG 802
Qy      201 AspProlylsleuCySProAlaAspProAspTrrPleAlaPheIleHisSerAsnAspIle 220
Db      803 GATCCAAAATTTATGCCCTGCTGATCCAGACTGATGCTTTTATACATACCAACGATATT 862
Qy      221 TrrPleSerAsnIleValThrArgGlnGlnuArgArgleuThrTyrsValHisAsnGluLeu 240
Db      863 TGGATTTCTTACATCGTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 922
Qy      241 AlaAsnMetGlnGlnuAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlnGlu 260
Db      923 GCCAACATGGAAGAGATGCCAGATCAGCTGAGTGCCTTACCTTTGTTCTCCAAAGAGAA 982
Qy      261 PheAspArgTyrsSerGlyTyrrTrrPrrCysProlylsAlaGluThrThrProSerGlyGly 280
Db      983 TTTGATAGAGATTTCTGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1042
Qy      281 LysIleLeuArgIleLeuTyrsGlnGlnuAsnAspGlnSerGlnuValGlnuIleHisVal 300
Db      1043 AAAATTTCTTGAATTTCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1102
Qy      301 ThrSerProMetLeuGlnuThrArgArgAlaAspSerPheArgTyrsProlysthrGlyThr 320
Db      1103 ACATCCCTCTATGTTGGAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1162
Qy      321 AlaAsnProlystValThrPheIleMetSerGlnuIleMetIleAspAlaGlnuIleArgIle 340
Db      1163 GCAAAATCTTAAAGTCACTTTAAGATGTCAGAAATATGATGATGTCGAGAGAGAGATC 1222
Qy      341 IleAspValIleAspIleGlnuLeuIleGlnProPheGlnuIleLeuPheGlnuIleValGln 360
Db      1223 ATTAGATGTCATAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1282
Qy      361 TyrrIleAlaArgAlaGlyTrrPrrProGlnuGlyLeuTyrsAlaTrrPserIleLeuLeuAsp 380
Db      1283 TATATTTGCCAGAGCTGGATGAGCTCTCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1342
Qy      381 ArgSerGlnThrArgleuGlnuIleValIleuIleSerProGlnuLeuPheIleProValGln 400
Db      1343 CGGTCCCAAGCTCGCTTACAGATAGTGTGATCTCACCTGAATTTATTTTCCAGAGAGAA 1402
Qy      401 AspAspValMetGlnuArgGlnuArgleuIleGlnuSerValProAspSerValThrProleu 420
Db      1403 GATGATGTTATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1462
Qy      421 IleIleTyrrGlnuIleThrThrAspIleTrrPleAsnIleHisAspIlePheHisValPhe 440
Db      1463 ATATCTATGAGAGAAACAACAGACATCTGGATTAATATCATGACATCTTTTCAATGTTTT 1522
Qy      441 ProGlnSerHisGlnuGlnuIleGlnuPheIlePheAlaSerGlnuCysIleThrGlyPhe 460
Db      1523 CCCCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1582
Qy      461 ArgHisLeuTyrrIleThrSerIleLeuIleGlnuSerTyrrTyrsArgSerSerGly 480

```

```

Db      1583 GGTCAATTTATACAAATTTATCATCTTATTTAAAGAGAAAGCAAAATATTAACATCCAGTGT 1642
Qy      481 GlyLeuProAlaProSerAspPheIleCysProIleTyrsGlnuIleAlaIleThrSer 500
Db      1643 GGGCTGCTCTGCTCCAAATGATTTTCAAGTCTCTATCAAAAGAGAGAGAGAGATTAACAGT 1702
Qy      501 GlyGlnuTrrPleValLeuGlnuIleArgHisGlySerAsnIleGlnuValAspGlnuValArg 520
Db      1703 GGTGAATGGAGAGTTCCTGGCCGAGATGATCTTAATATCCAAAGTTGATGAAGTCAGAGG 1762
Qy      521 LeuValTyrrPheGlnuIleThrIleAspSerProleuGlnuIleHisLeuTyrrValIleSer 540
Db      1763 CTGTATATTTTGAAGGACCAAGAGACTCCCTTTTAAAGATCACTGTACGTAGTACGT 1822
Qy      541 TyrrValAsnProGlnuIleValThrArgleuThrAspArgGlyTyrrSerHisSerCysCys 560
Db      1823 TACGTAAATCTCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1882
Qy      561 IleSerGlnHisCysAspPhePheIleSerTyrsSerAsnGlnuIleAsnProHisCys 580
Db      1883 ATCAGTCAGACAGTGTGCTTTTATAGATAGATAGTAAACAGAGAAATCCACAGCTGT 1942
Qy      581 ValSerLeuTyrrIleLeuSerSerSerProGlnuAspAspProThrCysIleThrIleGluPhe 600
Db      1943 GTGTCCCTTTACAGCTATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2002
Qy      601 TrrAlaThrIleLeuAspSerAlaGlyProleuProAspTyrrThrProProGlnuIlePhe 620
Db      2003 TGGGCCACCAATTTGATTTGACAGAGTCCCTTCTCTGATCTATCTCTCCAGAAATTTTC 2062
Qy      621 SerPheGlnuSerThrThrGlyPheThrLeuTyrrGlyMetLeuTyrrIleAspLeu 640
Db      2063 TCTTTTGAAGGATCTACTGATTTTACATGTATGGAGTGTCTTACAGAGCCATGATCTA 2122
Qy      641 GlnProGlnuIleTyrrIleTyrrProThrValLeuPheIleTyrrGlyGlyProGlnuValGln 660
Db      2123 CAGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2182
Qy      661 ValAsnAsnArgPheIleGlyValIleTyrrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db      2183 GTGAATATATGCTTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2242
Qy      681 TyrrValValValIleAspAsnArgGlySerCysHisArgGlyLeuIlePheGlnuIle 700
Db      2243 TATGTGTTGTAGTATGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2262
Qy      701 AlaPheIleTyrrIleMetGlyGlnuIleGlnuIleAspAspGlnuValGlnuIleGlnuIle 720
Db      2303 GCCTTTAAATATTAATGAGTCAAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2362
Qy      721 LeuAlaSerArgTyrrAspPheIleAspLeuAspArgValGlyIleHisGlyTrrPserTyrr 740
Db      2363 CTAGCTTCTCGATATGATTTTCAATGATCTTATGATGATGATGATGATGATGATGATGAT 2422
Qy      741 GlyGlyTyrrIleuSerLeuMetAlaLeuMetGlnuArgSerAspIlePheArgValAlaIle 760
Db      2423 GGAAGATACCTCTCCCTGATGAGCATTAATGACAGAGAGAGAGAGAGAGAGAGAGAGAG 2482
Qy      761 AlaGlyAlaProValThrIleuTrrPlePheTyrrAspThrGlyTyrrThrGlnuArgTyrrMet 780
Db      2483 GCTGGGAGCCCACTCACTCTGAGATCTTCTATGATACAGAGAGAGAGAGAGAGAGAGAG 2542
Qy      781 GlyHisProAspGlnuAsnGlnuIleGlyTyrrIleGlnuIleSerValAlaMetGlnuIleGln 800
Db      2543 GGTCAACCTCGACCAAGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2602
Qy      801 LysPheProSerGlnuProAsnArgleuLeuLeuIleGlyPheLeuAspGlnuAsnVal 820
Db      2603 AAGTTCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2662
Qy      821 HisPheAlaHisThrSerIleLeuLeuSerPheIleValArgAlaGlyLysProTyrrAsp 840
Db      2663 CATTTTGCACATACAGATATTTATGAGATTTTATGAGAGAGAGAGAGAGAGAGAGAGAGAG 2722

```


QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
 DB 1114 ACATCCCCCTATGTTGAAAACAAGAGGAGGACATTCCCTTATCTCTAAACAAGGTACA 1173
 QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluValArgIle 340
 DB 1174 GCAAATCTTAAGTCACTTTTAAAGATGCAAAATAATGATGATGCTGAAAGGAAGATC 1233
 QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleuPheGluValGlu 360
 DB 1234 ATAGATGTCATAGTAAAGAACTAATCAACCTTTTGAGATTCTATTGAAAGAGTTGAA 1293
 QY 361 TyrIleIleAsnArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIleLeuLeuAsp 380
 DB 1294 TATATTCACAGAGTGAATGAGATCTCGAGGGAAATAATGCTTGCTCATCTCACTGAT 1353
 QY 381 ArgSerGlnThrArgLeuGlnIleValIleuIleSerProGluLeuPheIleProValGlu 400
 DB 1354 CGCTCCAGACTCGCCCTACAGATGATGTTGATCTCACCTGAAATTAATTAATCCAGTAA 1413
 QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
 DB 1414 GATGATGTTATGAAAAGCAGAGACCTATTGAGTCAATGCTGATTTCTGAGCCGCTCA 1473
 QY 421 IleIleIleGlyGluGlnThrThrAspIleTyrPileAsnIleHisAspIlePheHisValPhe 440
 DB 1474 ATTATCTATGAAGAAACAACAAGACATCTGAAATAATTCATGATGATCATCTTCAATGTTT 1533
 QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
 DB 1534 CCCCAAGTCAACAAGAGGAATGAGTTGATTAATTTTCCCTGAAATCAAAACAGGTTTC 1593
 QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
 DB 1594 CGTCATTTATACAAATTAATCATCTATTTAAGGAAGCAAAATATAACATCAAGTCA 1653
 QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
 DB 1654 GGGCTCGCTGCTCAAGTGAATTCAGTGTCTCATCAAGAGAGATAGCAATTAACAGT 1713
 QY 501 GlyIleTyrProGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
 DB 1714 GGTGAATGGGAATCTTCTGCGCGGATGATCTTAATTCAGAGTTGATGAAAGCAGAGG 1773
 QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValIleSer 540
 DB 1774 CTGGTATATTGTAAGGCAACAAGACTCCCTTTAGACATCACTGTAAGTCACT 1833
 QY 541 TyrValAsnProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
 DB 1834 TACGTAATCTCTGAGAGGTTGACAGGCTGACTGACCGTGGCTACTACATCTTCTGCTGC 1893
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
 DB 1894 ATCAGTCACAGACGTGACTCTTTATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGT 1953
 QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
 DB 1954 GGTGCTCCCTTACAGACTATCAAGTCTGAAGATGACCAACTGCAAAACAAGGAATTT 2013
 QY 601 ThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
 DB 2014 TGGGCCCACATTTTGGATTCAGAGGTCCTCTTCCGACTATACCTCTCCAGAAATTTTC 2073
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGluMetLeuTyrLysProHisAspLeu 640
 DB 2074 TCTTTTGAAGAAGTACTGATTTACATCTGATGATGATGATGATGATGATGATGATGAT 2133
 QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
 DB 2134 CAGCCCTGGAAGAAATATCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2193
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680

DB 2194 GTGATATATCGGTTTAAAGAGTCAAGTATTTCCGCTTGATACCTTAGCTCTAGT 2253
 QY 681 TyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
 DB 2254 TATGTGGTTTATGATGATGACAAACAGGGATCCCTGATCCGAGGGCTTAAATTTGAAGGC 2313
 QY 701 AlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGlnValGluGlyLeuGlyTyr 720
 DB 2314 GCCTTTAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2373
 QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
 DB 2374 CTAGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2433
 QY 741 GlyIleTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
 DB 2434 GAGAGATACCTCTCCCTGATGACATTAATGACAGGTCAGATATCTTCAAGGTTGCTATT 2493
 QY 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
 DB 2494 GCTGGGCCCCAGTCACTCTGATGATTTCTTAATGATCAAGATACAGGAACTTATATG 2553
 QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
 DB 2554 GGTCACTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 2613
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
 DB 2614 AAGTTCCTCTTAAACCAATCGTTTAACTGCTTAACTGCTTAACTGCTTAACTGCTTAA 2673
 QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
 DB 2674 CATTTGCAATACCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2733
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
 DB 2734 TTACAGATCTATCTCTGAGAGACACAGCATTAAGTCTGATGATGAGAAATATAT 2793
 QY 861 GlnLeuHisIleLeuHisTyrTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
 DB 2794 GAATCGATCTTTTGAATCACTTCAAGAAACCTTGATGACGTTATGCTGCTCTTAA 2853
 QY 881 ValIle 882
 DB 2854 GTGATA 2859
 QY 882 ValIle 882
 DB 2854 GTGATA 2859

RESULT 6
 AX354793 3143 bp DNA linear PAT 06-FEB-2002
 LOCUS AX354793
 DEFINITION Sequence 1 from Patent WO0179473.
 ACCESSION AX354793
 VERSION AX354793.1 GI:18619526
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 9
 Meyers, R.A. and Williamson, M.
 21953, a human prolyl oligopeptidase family member and uses thereof
 Patent: WO 0179473-A 1 25-OCT-2001;
 Millennium Pharmaceuticals, Inc. (US)
 FEATURES
 source
 1..3143
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 229..2877
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD22990.1"
 /db_xref="GI:18619527"

/translation="MAAMETBQLGVEIFETADCEINTESODRPKLEPFYVERYSWQ
LKKLADTRKYHGYMAKAPHDPMFKNDPDGPHSDIYYILANGSENRENTLFPSEI
PKTINRAAVLMSKPLDLFOATLDYQWYSREBELERKRIIGVAGIADYHAGSG
TFLFOAGSGIYHVKGQPGQFTQCLPENVLSTSPNTRPMDCLCPAPDIAFHSHN
DIWISNITREERRITVYHNELIANEEDASAGVATPFIQSEFPDYSWYMCPRATT
DSGGKILFIYBENDESEVELIHVTSPLERBRADSPYKPTGTANPVTYTKMSIMI
DABGRIIDVIDELIOPELIFEGVEYIARAGMTPEGYASILLDRQTLQIVLIS
PELFIPEVDWNERQLIESVPDSVTPLIYEETDINIHIDIFHVPQSHIEIER
IFASECKTFPHLYKITSILKESKYSGGIIPASDPKCPKKEIALITSGMEVILGR
HGSNIQVDEVRLLVYFEGKDSPLEHLYVYSYVNGVTRLTDBYSGSCISQCHD
PFISKYSNOKPHCVSLKYKLSPEDDPKCKEPWATLIDSGPLPDYTPPEISFES
TTGPTLYGMLYKPHLDQPKKPYTLFIYGGEPQVQVUNNRKGYKPYPLANTLACIYV
VVVIDNRSGSRGKLFBSAFYKMGQIIDQVBSIQYLASRYDFTDIDRVGHSWY
GGYLSMALMQRSDFRVALIAGAPVLTMIFYDTQYTERMYGHPDNEGYIAGVMQ
AEKPESEPNRLILLHGPIDENVHFAHTSILLSFIVRAKPYDLQIYPOEHSIRVES
GEHYELHLHTLOENLSRIALAKVI"

BASE COUNT 943 a 644 c 712 g 844 t
ORIGIN

Alignment Scores:

Pred. No.: 0
Score: 4700.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%

Length: 3143
Matches: 882
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-070-464-1 (1-882) x AXJ54793 (1-3143)

QY 1 MetalAlaIaIaMetGluThrGluGluMetGluValGluIlePheGluThrAlaAspCys 20
DB 229 ATGGGAGCAGCATGGAACAGAACAGCTGGGTGTTGAGATATTTGAAACCTGGCAGCTG 288
QY 21 GluGluAsnIleGluSerGlnAspArgProIysLeuGluProPheTyrValGluArgTyr 40
DB 289 GAGGAGATATTGATTCACAGGATGGCTTAATTGGAGCCCTTTTATGTTGAGCGGTAAT 348
QY 41 SerTyrSerGlnLeuIleIysLeuLeuAlaAspThrArgIysTyrHisGlyTyrMetMet 60
DB 349 TCCTGGAGTCAGCTTAAAGCTGCTGCCGATACCAAGAAATATATCATGGCTACACTGATG 408
QY 61 AlaIysAlaProHisAspPheMetPheValIysArgAsnAspProAspGlyProHisSer 80
DB 409 GCTTAGGACCACTGATTTTCATGTTTGTGAAGAGAAATGATCCAGATGACCTTATTC 468
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyIleAsnArgIleAsnThrLeuPheTyrSer 100
DB 469 GACAGAACTCTATTACCTTGCCATGCTGCTGTGAGAACAGAAATACACTGTTTATTTCT 528
QY 101 GluIleProIysThrIleAsnArgAlaAlaValLeuMetLeuSerTyrIlyspProleuLeu 120
DB 529 GAAATTCCTCAAACTATCATATAGACACAGCTTAAATGCTCTCTTGGAAGCCTCTTTTG 588
QY 121 AspleuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluIleuLeuArg 140
DB 589 GATCTTTTTCAGGCAACCTGACCTATGGAATGATTTCTCGAAGAAAGAACTATTTAAGA 648
QY 141 GluArgIysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnIlyserGly 160
DB 649 GAAAGAAACCATGGAACAGCTCGAAATGCTCTTACGATTATCAACAAGAAAGTGA 708
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValIysAspGlyIleProGlnIly 180
DB 709 ACATTTCTGTTTCAAGCCGTAGTGAAATTTATTCAGTAAAGATGAGGGCCACAAGGA 768
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB 769 TTTAGGCAACAACCTTTAAGGCCCACTAGTGAAACTAGTTGCTCCAACTACGAGATG 828
QY 201 AspProIysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
DB 829 GATCCAAATATATGCTGCTGATCCAGACTGAGATTGCTTTTATATGATAGCAACGATATT 888

QY 221 TrpIleSerAsnIleValIleThrArgGluGluArgIleLeuThrTyrValHisAsnGluLeu 240
DB 889 TGATATCTTACATCGTATACCAAGAAAGAGAGACTCATATGTCACATAGAGCTA 948
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaIleThrPheValLeuGluGlu 260
DB 949 GCCAACATGGAAGAAAGAGCCAGATCAGCTGAGTGGCTACCTTGTTCTCCAAAGAA 1008
QY 261 PheAspArgTyrSerGlyTyrTyrTrpCysProIysAlaGluThrThrProSerGlyGly 280
DB 1009 TTTGATGATATTTCTGCTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1068
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleIysVal 300
DB 1069 AAAATCTTGAATCTCTATGAAAGAAATGATGATCTGAGGTGGAATATTATTCATGTT 1128
QY 306 ThrSerProMetLeuGluThrThrArgArgAlaAspSerPheArgTyrProIysThrGlyThr 320
DB 1129 ACATCCCTTATGTTGGAACAAAGAGGCGACATTCATTCGTTATCTTAAACAGGTACA 1188
QY 321 AlaAsnProIysValIleThrPheIysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
DB 1189 GCAATCTTAAAGTCACCTTTTAAGATGTCAGAAATATATGATGATGCTGAAAGAAAGATC 1248
QY 341 IleAspValIleAspIysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
DB 1249 ATAGATGCTATAGATTAAGAACTATATCACTTTTGAGATTCTATTTGAAAGAGATTGAA 1308
QY 361 TyrIleAlaArgAlaGlyTyrThrThrProGluIlyIysTyrAlaIleAspSerIleLeuLeuAsp 380
DB 1309 TATATGCGAAGCTGAGTATGAGATCTCTGAGGAAATATGCTGCTGATCATCTTAAAT 1368
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
DB 1369 CGCTCCCAAGATCGCTGCAATAGTGTGATCTTCACCTGAAATTTATTTATCCAGATAGA 1428
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
DB 1429 GATGATGTTATGAGAAAGGACAGACATCATGATGACAGTGGCTGATTCGTGACGCGACA 1488
QY 421 IleIleTyrGluGluThrThrAspIleThrIleAsnIleHisAspIlePheHisValPhe 440
DB 1489 ATATCTTATGAGAAACCAACAGACATCTGATTAATATCAATGATCTTTCATATGTTTTT 1548
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIysThrGlyPhe 460
DB 1549 CCCCAAGTCAAGAAAGGAATAGATTTATTTTGGCTCTGATGATGAAACAGGTTTC 1608
QY 461 ArgHisLeuTyrIlyIleThrSerIleLeuIysGluSerIysTyrIysArgSerSerGly 480
DB 1609 CGTCATTTATACAAATATACATCTATTTTAAAGAAAGCAATATTAACGATCCAGTGGT 1668
QY 481 GlyLeuProAlaProSerAspPheIysCysProIleIysGluGluIleAlaIleIleThrSer 500
DB 1668 GGGCTGCTGCTCCAGATGATTTCAAGGTCTCTATCCAAAGGAGATGACATTTACAGT 1728
QY 501 GlyIleTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
DB 1729 GGTGATGCGGAAGTCTTGCGCGGACATGATCTAATATCCAAAGTTGAGAAAGTCAAGAG 1788
QY 521 LeuValTyrPheGluGlyThrThrIysAspSerProLeuGluIleHisIleuTyrValIleSer 540
DB 1789 CTGGTATATTTTGAAGGACCAAAAGACTCCCTTTAGAGCATCACTTACGATGATGAGT 1848
QY 541 TyrValAsnProGlyIlyValIleThrArgLeuThrAspArgGlyTyrTyrSerHisSerCysCys 560
DB 1849 TACGTAATTCCTGAGAGGTGACAAAGGCTGACCTACCGTGGCTTCTCATCTTCTGCTGC 1908
QY 561 IleSerGlnHisCysAspPhePheIleSerIysTyrSerAsnGlnIlyAsnProHisCys 580
DB 1909 ATCAGTCAAGCACTGCTGATCTTTATAGTAAAGTATGATACCGAAAGATCCACACTGCT 1968
QY 581 ValSerLeuTyrIlyLeuSerSerProGluAspAspProThrCysIysThrIysGluPhe 600

```

Db      1969 GTGCCCCCTTACAGCTATCAAGTCCCTAAGATGACCCAACTGCGAAAACAAAGGAAATT 2028
Qy      601 TTPAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrTrpProGluIlePhe 620
Db      2029 TGGGCCACCATTTTGGATTGACAGAGTCTCTCTGCTGACTATCTCTCCAGAAATTTTC 2088
Qy      621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db      2089 TCTTTGAAAGTACTACTGAGATTACATGTATGGAGGCTCTCAAGGCTTCATGATCTA 2148
Qy      641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGluIleValGluLeu 660
Db      2149 CAGCTCGAAGAAATATCCATCTGCTGCTTCAATATGAGTGGCTTCAGAGTCCAGTTC 2208
Qy      661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db      2209 GTGAATATCGGTTTAAAGGAGCAAGTATTTCCGCTTGAATACCTTAGCTCTAGGT 2268
Qy      681 TyrValValValValIleAspAsnArgGlySerGlyHisArgGlyLeuLysPheGluGly 700
Db      2269 TATGTGGTGTAGTGAAGACAAACAGGAGATCTGTCAACGAGGCTTAAATTTGAAGGC 2328
Qy      701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
Db      2329 GCCTTTAAATATATAAATGGGTCAAAATGAAATTCAGATCAGGTGGAAGACTCCAAAT 2388
Qy      721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
Db      2389 CTAGCTCTCTGAAATGATTTCACTTGAATGCTGTGGGCAATCCAGGCTGGTCTCTAT 2448
Qy      741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db      2449 GGAGGAACTCTCTCCATGAGCTTAAAGCAGAGGTCAAGATCTTCAGGCTTGTCTATT 2508
Qy      761 AlaGlyAlaProValThrLeuTyrPheTyrAspThrGlyTyrThrGluArgTyrMet 780
Db      2509 GCTGGGGCCCCAGTCACTGCTGGATCTTCAATGATCAGAGATACAGGAACTTATATG 2568
Qy      781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
Db      2569 GGTCACTCCCTGACAGATGAAACAGGCTTACTTGAATCTGTGGCTGAGCAAGCAGAA 2628
Qy      801 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db      2629 AAGTTCCCTCTGAACCAATCGTTTACTGCTCTTACATGCTTCTGATGAGAAATGTC 2688
Qy      821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
Db      2689 CATTTGACATACCAAGTATATTACTGAGTTTTTGAAGGCTCGAAAGCCATATGAT 2748
Qy      841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
Db      2749 TTACAAATCTATCTCCTCAGAGACACAGCATTAAGATCTCTGAATCGGGAACAATTT 2808
Qy      861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
Db      2809 GAACCTCACTTTTGCACATCTTCAAGAAACCTTGATCACTGATGCTGCTCTTAA 2868
Qy      881 ValIle 882
Db      2869 GTGATA 2874

```

RESULT 7
 LOCUS BC040203 4535 bp mRNA linear PRI 03-DEC-2002
 DEFINITION Homo sapiens, similar to dipeptidyl peptidase 8, clone MGC:42570
 ACCESSION BC040203
 VERSION BC040203.1 GI:2607916
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE AUTHORS TITLE JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 4535)
 Strausberg, R.
 Direct Submission
 Submitted (27-NOV-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: gcgabs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louised, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nannavati,
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 71 Row: K Column: 23.

FEATURES source

1..4535
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:42570 IMAGE:4824813"
 /tissue_type="Testis"
 /clone_id="NH_MGC_97"
 /lab_host="DH10B"
 /note="Vector: pBluescript"
 1629..4277
 /codon_start=1
 /product="similar to dipeptidyl peptidase 8"
 /protein_id="AAH40203.1"
 /db_xref="GI:2607917"
 /translation="MAAAMEEQLGVEIFETADCEENISODRPKLEPFYVERYSWQ
 LKTLADTRKHYGMWAKA PHDMFVKRNDPOGPHDRIVYLMSGMRNTLFSRI
 PKTINRAALVLMSSKRLDLFOATLDYGMTSREELIRERKRTIGVIGASVDYHQS
 TTRPQSGGITHVKGDPQSTQDPLRPMLVETSCNIRNDPLCPADPWIAFIHGN
 DIWISIVTIREERLTYVNEELANMEEDARSAGVAFVLOEEDRSYGWVCKAEIT
 PSGGKILRIIEBENDESEVEIIVHTSPMLSTRADSFYKPTGAMPKVPKSEIMT
 DAGRIIDVIDKEILOPFEILFEGVEYIARAGTTPGKXAMSLILDRSOTRLOIUIS
 PELFIPVEDDVERORLIRSVDSVPLIIVERTTQWIMINHIIFHVPOSHHEIEF
 TRASECKTGRPHLYKTTSLIKESKRGSSGGLPAPDFKPIIEEITATISGEVEIEF
 HSNIVDVERLVFEGTKDPLEBHLIVSVSVNGEVRLTDROYSHSCCISOHD
 FPISTKSNQKNHCVSLYKLSPEDEPTKTEFMATILDSAGLPDYPPELFSFES
 TTGPTLYGMILYRPHDLOPKKTPVLFIVGPOVOLLNRRFKVYFRNLTLASLGY
 VVIVINRGSCHRGLEKFGAKFYKMGQIIEIDOVGQYASRDFIDLRVGHGYSY
 GGLYSIMALMORSDIPIRAIAGAPVLTATFVDYGTERTYRWGHPDQEGCYVGSVMO
 AKRPSSEPRILLIGFDENVRPAFTSLILSFLVAKGPRYDQLQITPQERHSTRVES
 GEHYELHLIHTYQENUGSRITAKVT"

CDS

BASE COUNT 1313 a 918 c 946 g 1358 t
 ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4535
 Score: 4696.00 Matches: 881
 Percent Similarity: 99.89% Conservative: 0
 Best Local Similarity: 99.89% Mismatches: 1
 Query Match: 99.91% Indels: 0
 DB: 0 Gaps: 0

US-10-070-464-1 (1-882) x BC040203 (1-4535)

```
QY      1 MetalalaIaIaMeGIuThrGIuGIuMeGIuValGIuIlePheGIuThrAlaSPCyS 20
Db      1629 ATGCAGACGACNAATGAAACAGAACAGCTGGGTGTGGAATATTGAAACCTGGCAGCTGT 1688
QY      21 GluGIuAaenIIeGIuSerGIuAaSPArProIySLeuGIuProPheTyValGIuArgTy 40
Db      1689 GAGGGAATATTGAATCAACAGATCGGCTTAATTGAGCCCTTTTATGTTGAGCGGTAT 1748
QY      41 SerTPSerGIuLeuIySLeuLeuAlaSPThrArgIySTyRHSIGIYTyRMeMet 60
Db      1749 TCTTGAGATCAGCTTAAAAAGCTGCTTCCCATACCAAAATATCATCGCTACATGATG 1808
QY      61 AlaIySaIaProHISaSPPheMeTPheValIySaArgAaSPProaSPGIProHISer 80
Db      1809 GCTAAGGACACACATGATTTCACTGTTGTGAAGAGAAATGATCCAGATGACCTCATTC 1868
QY      81 AsPaArgIleTyTyLeuAlaMeSerGIyGIuAaenArgIuAaenThrLeuPheTySer 100
Db      1869 GACGAATCTATTAATCTTGCCATGCTGTGTGAAGACAGAAATACACTGTTTATTTCT 1928
QY      101 GluIleProIySThrIleAaSPArgAlaAlaValIeMeTLeuSerTPlySPProleu 120
Db      1929 GAAATTCCTCAAAACATATCAATAGACACAGCTTAATGCTCTCTTGAGAACCTCTTTG 1988
QY      121 AsPleuPheGIuAlaThrLeuAaSPTyGIyMeTySerArgGIuGIuLeuLeuArg 140
Db      1989 GATCTTTTTCAGGACACATGACTATGAAATGATCTTCGAGAAAGAACTATTAGA 2048
QY      141 GluArgIySaArgIleGIYThrValGIYIleAlaSerTyRaSPTyRHSIGIYSerGI 160
Db      2049 GAAAGAAACCATTTGGAACAGTCGGAATTCCTTTCGATTAACCAAGAAAGTGG 2108
QY      161 ThrPheLeuPheGIuAlaGIYSerGIYIleTyRHSaValIySaSPGIYGIProGIu 180
Db      2109 ACATTTCTGTTTCAAGCCGATAGTGAATTTATCACTGAATAAGATGAGGGCCACAAGA 2168
QY      181 PheThrGIuGIuProleuArgProAaenIleuValGIuThrSerCySPProAaenIIeaRgMet 200
Db      2169 TTTAGCAACACACTTTAAGGCCCAATCTAGTGAAGAACTAGTTGCCCAACATACGGATG 2228
QY      201 AsPProIySLeuCySPProAlaSPProaSPTrpIleAlaPheIleHISerAaSPle 220
Db      2229 GATCCAAATTAATGATGCCCTGCTGATCCAGACTGGAATGCTTTTATACATGGAACGATATT 2288
QY      221 TrpIleSerAaenIIeValIThrArgGIuGIuArgLeuThrTyValHISaSPGIuLeu 240
Db      2289 TGGATATCTAATCATCTGAACAGAGAAAGAAAGAGACTCACTTAATGTCACAATGAGCTA 2348
QY      241 AlaAaenMeGIuGIuAaSPAlaArgSerAlaGIYValAlaThrPheValIeGIuGIu 260
Db      2349 GCCAACATGGAAGAGATGCCAGATCACTGAGTGGCTACCTTTGTTCTCCAGAAAGAA 2408
QY      261 PheAaSPArgIySerGIYTyRTrpTyRProIySaIaGIuThrThrProSerGIYGIY 280
Db      2409 TTTGATGATATTTCGGCTATTGGGTGGTGTCCAAAGCTGAACACTCCCAAGTGGTGT 2468
QY      281 LysIleIleuAaGIleuTyRGIuGIuAaAaSPGIuSerGIuValGIuIleIleHISaVal 300
Db      2469 AAAATTTCTTAAGATTCATATGAAAGAAATGATGATCTGAGTGGGAATTAATTCATGTT 2528
QY      301 ThSerProMeTLeuGIuThrArgArgAlaSPSerPheArgTyRProIySThrGIYThr 320
Db      2529 ACAATCCCTATGTTGGAACAAAGAGGCAATTCATTCCTTAATCCTAAGAAACGGTACA 2588
QY      321 AlaAaenProIySaIThrPheIyMeSerGIuIleMeTleAaSPaGIuGIuIleArgIle 340
Db      2589 GCAATTCCTTAAGACACTTTTAAGATGTCAGAAATTAATGATGATGTCGAAAGAAAGATC 2648
QY      341 IleSPaValIleAaSPySGIuLeuIleGIuProPheGIuIleLeuPheGIuGIuValGIu 360
Db      2649 ATAATGTCATAGTAAGAACTAATTCACCTTTTGAGATTCATATTGGAAGGAGTTGAA 2708
QY      361 TyRleAlaIaArgAlaGIYTrpThrProGIuGIuIySTyRAlaTrpSerIleLeuLeuAaSP 380
Db      2709 TATATGGCAGAGCTGGATGGAATCTCGTAGGGAATAATGCTGTGCTCATCTCATATAT 2768
QY      381 ArgSerGIuThrArgLeuGIuIleValIleuIleSerProGIuLeuPheIleProValGIu 400
Db      2769 CGCTCCACAGACTCCCTACAGATAGTGTGATTCACCTCGAATTAATTAATCCAGTAGAA 2828
QY      401 AsPaSPaValMeGIuArgGIuArgLeuIleGIuSerValIProaSPSerValIThrProleu 420
Db      2829 GATGATGTTATGAAAGGACAGAGACTCATTAAGTGAAGTCCGATTCGTGAGCGCCACTA 2888
QY      421 IleIleTyRGIuGIuThrThrAspIleTrpIleAaenIIeHISaSPIlePheHISaValPhe 440
Db      2889 ATTATCTATGAAGAAACAAACAGACATCGAATTAATATCCATGACATCTTTCACTTTT 2948
QY      441 ProGIuSerHISIGIuGIuGIuIleGIuPheIlePheAlaSerGIuCyLeuSThrGIYPhe 460
Db      2949 CCCCAGAGTCACGAAGAGAAATGAGTTATTTTGTGCTCGAATGCAAAAACAGATTTTC 3008
QY      461 ArgHISLeuTyRISIIeThrSerIleuLeuIySGIuSerIySTyRAsArgSerGIY 480
Db      3009 CCTCATTTATACAAATTAATCACTTAATTTTAAAGAAACCAATATAACATCCAGTGGT 3068
QY      481 GIYLeuProAlaProSerAaSPPheIyCySPProIleIySGIuGIuIleAlaIleThrSer 500
Db      3069 GAGCTGCTGCTCCAGAGATTTCAAGTGTCTATCAAAAGAGAGATGCAATTACAGT 3128
QY      501 GIYGIuTrpGIuValIeGIuArgHISGIYSerAaenIIeGIuValaSPGIuValArgTy 520
Db      3129 GGTAAAGTGAAGATTTCTTGGCCGATGGAATTAATCCAAAGTTGAAGAGTCAAGAG 3188
QY      521 LeuValTyRPhGIuGIYThrIyAaSPSerProleuGIuHISISLeuTyValIaSer 540
Db      3189 CTGCTATATATTTGAAGGACCAAAAGACTCCCTTTAGAGCATCACTGTACGTACTGAGT 3248
QY      541 TyRValaAaenProGIyGIuValIThrArgLeuThrAsPaArgIYTySerHISerCyS 560
Db      3249 TACGTAAATCCTGAGAGAGTGAACAAGCTGACTGACCGTGGCTACATCTTCTTGCTGC 3308
QY      561 IleSerGIuHISCySaSPPhePheIleSerIySTySerAaenIlySaSPProHISyS 580
Db      3309 ATCAGTACGACCTGTGACTTCTTTATGAAGTAATGATTAACCAAGAAATCCACACAGT 3368
QY      581 ValSerLeuTyRILySLeuSerSerProGIuAaSPProTrpCySLySThrGIYGIuPhe 600
Db      3369 GTGTCCTTTACAGACATATCAAGTCCGAGATGATCCCACTTGCAGAAACAAAGAAATT 3428
QY      601 TrpAlaThrIleLeuAaSPSerAlaGIYProleuProaSPTyRThrProProGIuIlePhe 620
Db      3429 TGGGCACACATTTTGGATTGACGAGTCCCTTCTGACTATACCTCCAGAAATTTTC 3488
QY      621 SerPheGIuSerThrThrGIYPhEThrLeuTyRGIYMeTLeuTyRILySPProHISaSPleu 640
Db      3489 TCTTTTAAAGTACTTACGATTTACATGTATGAGAGTCTTACAGAGCTCCAGATCTA 3548
QY      641 GIuProGIyIySLySTyRProThrValIleuPheIleTyRGIYGIYProGIuValaGIuLeu 660
Db      3549 CAGCTCGAAAGAAATATTCCTACTGCTGTCTATATATGTTGCTCTCAGAGTCAAGTGTG 3608
QY      661 ValaAaAaArgPheIySGIYValIySTyRPhArgLeuAaenThrLeuAlaSerLeuGIY 680
Db      3609 GTGAATATCGGTTTAAAGAGTCACAGTATTTCCGTGTAATACCTCAGGCTCCTCTAGT 3668
QY      681 TyRValaValaValIleAaSPaAaArgGIYSerCySHISaRGIYLeuLySPPhGIuGIY 700
Db      3669 TATGTGTTGTAGTGAAGACCAAGGGGATCTCTGTACCGAGGCTTAAATTTGAAGGC 3728
QY      701 AlaPheIySTyRILySMeTGIYGIuIleGIuIleAaSPaSPGIuValaGIuGIuIySTyR 720
Db      3729 GCCTTTAAATATAAATAGGCTCAAAATGAATATGACATCAGGTGGAAGGCTCAATAT 3788
QY      721 LeuAlaSerArgTyRSPheIleAaSPleAaArgValaGIYIleHISGIYTrpSerTyR 740
```


Db 1354 CGCTCCAGACTCGCCCTACAGATAGTGTGATCTCACTGAATTAATTCACAGTAGA 1413
 Qy AspaSpValMetGluArgGlnArgLeuIleGlnSerValProAspSerValThrProLeu 420
 Db 1414 GATGATGTTATGGAAGGAGAGACTCATTGAGTCAGTCCGATCTCTGTGACGCACTA 1473
 Qy 421 ILeIeTYrGluGlnThrThraSpIleTyrIleasnIleHisAspIlePheHisValPhe 440
 Db 1474 ATTACTCTAGAGAAACAACAGACATCTGATTAATATCCATGATCATCTTTTCACTTTT 1533
 Qy 441 ProGlnSerHisGluGlnGlnIleGlnPheIlePheIleAspGlnCysAlaSerThrGlyPhe 460
 Db 1534 CCCCAAGTCACGAAAGAGAAATTAAGTTATTTTGTGCTGGAATGCAAAAACAGGTTTC 1593
 Qy 461 ArgHisLeuTYrIleTYrSerIleLeuLeuGlnSerIleTYrTYrLeuArgSerSerGly 480
 Db 1594 CGTCATTTATACAAATTAATTCATTTTAAAGAAACAAATTAATTAACGATCCAGTGGT 1653
 Qy 481 GlyLeuProAlaProSerAspPheIleCysProIleGlyGlnIleAlaIleThrSer 500
 Db 1654 GGGCGCTGCTCCAGATGATTTCAAGTGTCTCTATCAAAAGAGATAGCAATTCAGT 1713
 Qy 501 GlyGlnTTPGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
 Db 1714 GGTGAATGGAAAGTTCTTGCGCGGCAAGATTAATTCAGATTGATGAAAGTCAGAAAG 1773
 Qy 521 LeuValTYrPheGluGlyThyIleAspSerProLeuGlnHisHisLeuTYrValValSer 540
 Db 1774 CTGGTAATTTTGAAGGACCAAGACATCCCTTTAGAGCAATCACTGATGATGATGAT 1833
 Qy 541 TyrValAspProGlyGluValThraGlnThrArgLeuThrAspArgGlyTYrSerHisSerCys 560
 Db 1834 TACGTAATCTCGAGAGAGTACAGAGCTGACCTGACCTGCTGCTGCTGCTGCTGCTG 1893
 Qy 561 ILeSerGlnHisCysAspPhePheIleSerIleTYrSerAsnGlnIleAspProHisCys 580
 Db 1894 ATCAGTACGACACTGAGCTTCTTTATAGTAAGTATTAATTAACAGAGAAATCCACACTGT 1953
 Qy 581 ValSerLeuTYrIleLeuSerSerProGlnAspAspProThrCysAlaSerThrIleGlnPhe 600
 Db 1954 GTGTCCTCTTACAAAGCTATCAAGTCTGAGATGACCCACCTTGCAAAACAAAGAAATTT 2013
 Qy 601 ThrIleThrIleLeuAspSerAlaGlyProLeuProAspTYrThrProProGlnIlePhe 620
 Db 2014 TGGGCGACATTTTGATTCAGCGAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2073
 Qy 621 SerPheGlnSerThrThrGlyPheThrLeuTYrGlyMetLeuTYrIleProHisAspLeu 640
 Db 2074 TCTTTTGAAGTACTACAGTGAATTAACATTTGATGGAGCTCTTACAGCCCTCAATGATTA 2133
 Qy 641 GlnProGlyIleIleValTYrProThrValLeuPheIleTYrGlyGly- ProGlnValGln 660
 Db 2134 CAGCCTGGAAGAAATATCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2193
 Qy 660 euValAsnAspArgPheIleGlyValIleTYrPheArgLeuAsnThrLeuIleAsnLeu 680
 Db 2194 TGGGTAATTCAGGTTTAAAGAGTCAAGATTTCCGTTAAATTAACCTTACCTCTCTAG 2253
 Qy 680 ILYTYrValIleValIleAspAsnArgIleCysHisArgGlyLeuIleAspPheGln 700
 Db 2254 GTATATGTTGTGTGTGATAGCAACAGCGGATCTCTGCAACGAGGCTTAAATTTGAG 2313
 Qy 700 IYAlaIlePheIleTYrIleMetGlyGlnIleGlnIleAspAspGlnValGlnIleGln 720
 Db 2314 GCGCCTTTAAATATAAATGAGTCAAAATGAGAAATTCAGATCAGTGAAGAGATCCAAAT 2373
 Qy 720 YrLeuAlaSerArgTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTYrThrSer 740
 Db 2374 ATCTAGCTTCTCGATATGATTTTCACTTGAATCTGCTGCGGATTCACGCGTGGTCT 2433
 Qy 740 YrGlyGlyTYrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaI 760
 Db 2434 ATGAGAGATACCTCTCCCTGATGATTAATGACAGAGGTGATATATCTTCCAGGTTCTCTA 2493

Qy 769 IeAlaGlyAlaProValThrLeuTyrPheIlePheTYrAspThrGlyTYrThrGlnArgTYr 780
 Db 2494 TTGCTGGGAGCCCACTGCTCTGTGATCTTCTTATGATACAGATACAGGAACTTATA 2553
 Qy 780 etGlyHisProAspGlnAsnGlnGlnIleTYrTYrLeuGlySerValAlaMetGlnAla 800
 Db 2554 TGGGTACCTCGACCAAGATGAACAGGCTTTACTTATGATTCGTGGCCATTCAGACAG 2613
 Qy 800 IYIlePheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsn 820
 Db 2614 AAAAGTTCCCTCGAACAATGCTTTTACGCTCTTCAATGATGTTCTCGATGAGAAATG 2673
 Qy 820 aHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIleProTYr 840
 Db 2674 TCCATTTTGCACATACAGATATATTAATTAAGATTTTATGAGAGGCTGAAAGCCATATG 2733
 Qy 840 sPLeuGlnIleTYrProGlnIleArgHisSerIleArgValProGlnSerGlyGlnIleST 860
 Db 2734 ATTATACAGATCTATCTTCAGAGAGACACAGCATTAAGATTCCTGAATCGGAGAACAT 2793
 Qy 860 YrGluLeuHisLeuLeuHisTYrIleGlnGlnIleLeuLeuLeuGlySerArgIleAlaLeu 880
 Db 2794 ATGAACTGCATCTTTTGTGATTAACCTTCAAGAAACCTTGATCAAGATTCGCTCTTA 2853
 Qy 880 YsValIle 882
 Db 2854 AAGTGAT 2861
 RESULT 9
 BC043124
 LOCUS
 DEFINITION Mus musculus dipeptidylpeptidase 8, mRNA (cdna clone MGC:58057
 IMAGE:6410075), complete cds.
 ACCESSION BC043124
 VERSION BC043124.1 GI:27695449
 KEYWORDS MGC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 REFERENCE
 AUTHORS Dukayova, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 4799)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Dergs, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, K.H., Scheiner, C.F., Bhat, N.K.,
 Altschuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, T.L.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, J.A., Rubin, G.M., Hong, L.,
 Scalet, M., Soares, M.B., Donald, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loggellano, N.A., Peters, G.J.,
 Abramson, R.D., Mulhany, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.D., Hult, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.T., Skalska, U., Smalys, D.E.,
 Schermer, A., Schein, J.E., Jones, S.V., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 JOURNAL MEDLINE
 MEDLINE 2388257
 PUBMED 12477932
 REFERENCE
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-JAN-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

QY 400 GUAASPVALMERGLUARGGLNARGLEULIEGLUSERVALPROASPSERVALTHRPRO 419
 DB 1515 GAAGATGATGAGCCATGACAGACAGACTTATAGATGAGTCTGACTGCTGACACCA 1574
 QY 420 LEULIELEIETRYGLUGLUNTHRTHRASPILLETPILEASPILEHISAPPILEHEHISVAL 439
 DB 1575 CTGATCATCTATGAAGAAACAACAAGACATCTGATTAATATCCAGATATTTTTCATGTT 1634
 QY 440 PHEPROGLINSEHISGLUGLUNLIEGLUPHEILEPHEALASERGLUCYSIYSTRHGLY 459
 DB 1635 TTCTCTCAACATCAGAGATGAATGAAATTGATTATTTTGGCTCTGATGCAAAACAGGT 1694
 QY 460 PHEAIGHISLEUTRYLISYLETHRISERILEULYSGIUSERLYSTRYLYSARGSESER 479
 DB 1695 TTTCCATCTCTGATTAATATCAATCATCTTTAAAGAGAGCAATATATTAACGTCAGT 1754
 QY 480 GLIYGLIYLEUPROAIPROSERASPHELYSCYSPROILEYSGIUNLIEALIELETHR 499
 DB 1755 GGTGACATCACTGCCCCCAAGATATTCAGATGCTCATCAAGAAAGAAATTAACAATTACC 1814
 QY 500 SERGIYIUTRPGIUVALLIENGLIYARGHISGLISERASNIIEGLINVALASPGIUVALLIY 519
 DB 1815 AGTGGTGAATGGGAAGTACTTGCCCGCATGATCTAATATCTGGGTTGATGAAACCCAGA 1874
 QY 520 ARGLEUVALIYTRPHEGLUGLYTHIRIYSAASPSERPROLEUGIUNHISILEUTRYVALI 539
 DB 1875 AAGCTGGTGTACTTTGAAAGGACCAAGACCTCTCTTTGGAGCATCACCTGACCTGAC 1934
 QY 540 SERTYVALIASPPIROGLIYUVALITHRARGLEUTHRASPARGIYSTRYSEHISERYS 559
 DB 1935 AGTATATGAACACCTGGGAAGTGTGAGGCTGACTGACCGTGGTACTTCACACCTCTGC 1994
 QY 560 CYSIISESGIINHISCYSAASPSHEPHEILESERIYSTRYSERASNIIEGLINVALASNP 579
 DB 1995 TGCCCTGACCCGCGCATTTGACTCTTCTCATAGTATGACAGAACCAAGATCCACAC 2054
 QY 580 CYSVALISERLEUTRYLYLEUSERSERPROGLIUSPAPPROTHRCYSLYSTRYGLI 599
 DB 2055 TGTGTGTCCTCTCAAACTCTCAAGTCTCAAGATGACCAAGTTCATTAACAAAGGAA 2114
 QY 600 PHEITPALATHRIIELEUSASERIALGIYPROLEUPROAPPIYTRHTRPROGLIUNLIE 619
 DB 2115 TTTTGGGCGCACATTTTGGATTTCAGACAGTCTCTCTTGACTGACACCCCTCCAGAAATT 2174
 QY 620 PHESERPHEGLIUSERTHRTHIRGLYPHETHRLEUTRYGLIMETLEUTRYLYSPROHISASP 639
 DB 2175 TTTTCTTTGAAAGTACTGACTGATTTACACTGATGAGATGTTGATTAACCTCATGAC 2234
 QY 640 LEUGIINPROGLIYLYSILEUTRYPROTHRVALLIENPHEILEIETRYGLIYGLYPROGLIUNLIE 659
 DB 2235 CTACACCTGGAAGAAATACCCCACTGTGTTATTCATATVGTGTGTCCTCCAGAGTGCAG 2294
 QY 660 LEUVALIASNAANARGPHELYSGIYVALIYSTRYPHEARGLEUASNTHRLEUALISERLEU 679
 DB 2295 CTGGTGAACATCGGTTTAAAGAGTCAAGATTTCCGCCGGAACACCCCTGCCCTCCCTG 2354
 QY 680 GLIYTRYVALIYVALIYVALIYLEASPSANARGIYSECYSHISARGIYLEULYSPHEGLI 699
 DB 2355 GGTATGTGTGTGGTGGATGACACAGGGGATCTGTCAACGAGACCTTAATTGAA 2414
 QY 700 GLIYALIPHELYSTRYLSMERGLIYGLIUNLIEGLIUNLIEASPSAPGIUNVALIYGLIUNLIE 719
 DB 2415 GGGCCCTTTAAATTTAAATGAGGTGCATATAGAAATTCATATGATGAGGAGGAGCTCAG 2474
 QY 720 TYRLEUALISERARGIYRASPHEILEASPLEUSAPARGVALIYIIEHISGLYTRIPSER 739
 DB 2475 TACCTACATCTCAAGTATGACTTCATGACTTGATGAGAGTGGGCACTCAACGCTGTGTTCC 2534
 QY 740 TYRGLIYGLIYTRYLEUSERLEUMETALALEUMETGLINRGERASPIIEPHEARGVALIY 759
 DB 2535 TATGTGTGCTACCTCTCCGTGATGAGCATATATGACAGAGTGTGATATCTTCGGGTGCT 2594

QY 760 ILEALAGIYALIPROVALTHREUTRPILEPHERYASPTNRGLYTRYTHRGLIUNARGIY 779
 DB 2595 ATTGCTGGGGCCCCGATCACCTCGTGGATCTTTATGATATACAGATATACAGGAGGCTAT 2654
 QY 780 METGLIYISPROASPSIANSGLIUNGLIYTRYTRYLEUGIYSEVALIALAMETGLINAL 799
 DB 2655 ATGGGTCACTCCGACCAAGATGAACAGAGGCTACTATAGATCTGTGGCCATGCACAGCG 2714
 QY 800 GLIYSPHEPROSEGLIUPROASNARGLEULEULEUHIISGLYPHELEUASPGIUNSN 819
 DB 2715 GAGAGTTCCTCCCTGAAACCAACCGGTACTCTCTTGCAATGGGTTCTTGATATAGAT 2774
 QY 820 VALHISPEALAHISSTRSERILEULEUSERPHELEUVALIARGALIGLYSPROTRY 839
 DB 2775 GTTCACTTGCACACACAGATATATGCTGACTTTTATAGTGAAGGCTGGAAAGCATAT 2834
 QY 840 ASPLEUINLIEYTRPROGLIUNARGHISERILEARGVALIYPROGLISERGIYGLIUNHIS 859
 DB 2835 GACTTACAGATCTATCTCAGAGAGGACAGCATCAGAGTTCTCGAGTCTGAGAACAC 2894
 QY 860 TYRGLIUNLIEHISLEULEUHIISTRYLEUGIUNLIEUGIYSEYARGIILEALIALEU 879
 DB 2895 TATGACTGACACCTGCTCCTCACTTCAAGAGAACCTTGAATGCGATGCTGCTGTG 2954
 QY 880 LYSVALILE 882
 DB 2955 AAAGTGATA 2963
 RESULT 10
 LOCUS BC030688
 DEFINITION Homo sapiens, Similar to dipeptidylpeptidase 8, clone MGC:26191
 ACCESSION BC030688
 VERSION BC030688.1 GI:21265132
 KEYWORDS MGC.
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 3130)
 AUTHORS Strausberg,R.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAY-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NHI-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Miklos Palkevics, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Keteleman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whitting
 Clome distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 33 Row: d Column: 5
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 18450277.
 FEATURES
 source
 1..3130
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:26191 IMAGE:4822550"
 /tissue_type="Testis"
 /clone_lib="NHI_MGC_97"

CDS

/lab_host="DH00B"
/note="Vector: pBluescript"
372.2867
/codon_start=1
/product="Similar to dipeptidylpeptidase 8"
/protein_id="AAH0688.1"
/db_xref="gi:21265133"

/translation="MAAAMETQUGVEFETADCEBENTESDRPEFVYRYSWQ
LKULADRTKHYMMAPDHFVFNKNDGPHSRIRIYLAISGRENTEFLYSEI
PTFNRAAVLMSKPLDLFDLQATLDYDMYREBELRGRIGITGVASDYHSG
TFLEAGSGLVHVDGQFQOPLRNLYVETSOPNIRMPKICPADPKIATIHNS
DIWISNITREERLTYYHNEAMBERASGATATVLOEPRRYSYGMWCPRAET
PSGGKILRLIENEDESVEIHTYSPLETRRADSPRYPTGTANPKVTKSEIM
DABRRIIDVDKELIOPPEILFEGEYIARAGWTEGKAWMSILLDSQRTLIYLS
PELTPVEDDWERQRLIESVDSVTPLIYEETDIDINIHDFHVFPOSEIEIF
IFASECKTFRHLTKITSLKESYKRSGLPAPDFKPIKEEIAITSGMEVLSR
HGSNIQVDEVRILYFEGTKDSPLEHLVYVSYPGEVTRPLDGYSHSCISQHD
FPIKSYNOKNPHCVSLYKLSPPEDPCKEKFWATILDSAGPLDPYTPETISPS
TTGFTLGMLYKPHDLPQKKYPTVLFYKGPQYQVNNRKYKTRFLNLTASLGY
VVYTDNGSCHRGKKEGAPRYKVAIAGAVTLMITFYDGYTERYGHDPNOGY
LGSYVADKEKPSPEPNRLHLHGLDENVAHFATSIILSLVRAGEYDILQYQERN
SIRVPSGEHYEHLHLYQENLGSRTAAKVI"

BASE COUNT 955 a 626 c 717 g 832 t

Alignment Scores:

Pred. No.: 0 Length: 3130
Score: 4397.50 Matches: 831
Percent Similarity: 94.22% Conservative: 0
Best Local Similarity: 94.22% Mismatches: 0
Query Match: 93.56% Indels: 51
DB: 9 Gaps: 1

US-10-070-464-1 (1-882) x BC030688 (1-3130)

QY 1 MetAlaAlaIaIaMeGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
DB 372 ATGGCAGCAGCAATGAGAAACAGAACACACTGGGTGATTAATTTGAACCTGGCGACTGT 431
QY 21 GluGluAsnIleGluSerGlnAspArgProIysLeuGlnProPheTyrValGluArgTyr 40
DB 432 GAGGAGAAATATGATACAGAGATCGGCTTAATTTGAGCCTTTTATGTTGACGGTAT 491
QY 41 SerTrpSerGlnLeuIysIleuLeuAlaAspThrArgIysTyrHisGlyTyrMetMet 60
DB 492 TCCTGAGTACGCTTAAAGAGCTGCTGGCATACCGAATAATCATGCTCATCATGATG 551
QY 61 AlaIysAlaProHisAspPheMetPheValIysArgAsnAspProAspGlyProHisSer 80
DB 552 GCTAAGGACACACATGATTTCACTGTTGTGAAGAGAAATGATCCAGATGAGACCTCATTC 611
QY 81 AspArgIleTyrTyrIleuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
DB 612 GACAGATCTATTAACCTTCCATGCTGGTGAAGAGAACAGAAATATCATCTGTTTATTC 671
QY 101 GluIleProIysThrIleAsnArgAlaIleValIleuMetLeuSerTrpIysProLeuLeu 120
DB 672 GAAATTCCTAAACCTATCATAGAGAGACGCTTAATGCTCTCTTGAAGACCTCTTTTG 721
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
DB 732 GATCTTTTCAAGCACAACCTGACATGGAATGATATTCGAGAGAAAGACATTAAGA 791
QY 141 GluArgIysArgIleGlyThrValGlyIleAspTyrAspTyrHisGlnIysSerGly 160
DB 792 GAAAGAAAACGATTCGAGACAGTCGGAATGCTTCAAGATTATCATCAAGGAAGTGG 851
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValIleAspGlyGlyProGlnGly 180
DB 852 AATTTCTGTTTCAACCCGATGAGAAATTTATCATGTAAGAAAGTGAAGGACCAAGGA 911
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200

DB 912 TTACGCAACACCTTTAAGGCCCAATCTAGTGAAGAACTAGTGTCCCAACATACGATG 971
QY 201 AspProIysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
DB 972 GATCCAAATATATGCGCTGATCCAGACCTGAGATGCTTTTATACATAGCAACATATTT 1031
QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
DB 1092 TGGATATCTAATCATGTAACCAAGAGAAAGAGACCTCATTAATGTGCACATATAGCTA 1091
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaIleThrPheValIleuGluGlu 260
DB 1092 GCCAACATGGAAGAAATGCGACAGATCACTGAGTCCGCTACCTTGTCTCCAAAGAGA 1151
QY 261 PheAspArgTyrSerGlyTyrTrpTrpCysProIysAlaGluThrThrProSerGlyGly 280
DB 1152 TTTATAGATATATCTGGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1211
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
DB 1212 AAAATTTCTTGAATTTCTATATGAGAAATATGATATCTGAGGTGGAATTTATCATGTT 1271
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProIysThrGlyThr 320
DB 1272 ACATCCCTATGTTGGAACAGAGAGGAGGAGATTCATTCCTGATCTTAAACAGGTACA 1331
QY 329 AlaAspProIysValThrPheIysMetSerGlnIleMetIleAspAlaGluGlyArgIle 340
DB 1332 GCAATATCTTAATGATCATTTTAAATATGATATGATGATGATGATGATGATGATGATG 1391
QY 341 IleAspValIleAspIysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
DB 1392 ATAGATGTCATAGTAAAGAACTAATTCACCTTTTATGATTCATTTTGAAGAGCTTGA 1451
QY 361 TyrIleAlaArgAlaGlyTyrThrProGluGlyIysTyrAlaTrpSerIleLeuLeuAsp 380
DB 1452 TATATGCGCAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1511
QY 381 ArgSerGlnThrArgLeuGlnIleValIleIleSerProGluLeuPheIleProValGlu 400
DB 1512 CGCTCCACAGCTGACCTTACAGATGATGATGATGATGATGATGATGATGATGATGATG 1571
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
DB 1572 GATGATGTTATGGAAGGAGAGACATCATTAATGATGATGATGATGATGATGATGATG 1631
QY 421 IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
DB 1632 ATTATCTATGAAAGAAACAACAGACATCTGATTAATATCATGATCATCTTCAATGTTT 1691
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheIleAspGluCysIysThrGlyPhe 460
DB 1692 CCCCAAGTACAGAAAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1751
QY 461 ArgHisLeuTyrIleTyrSerIleLeuIysGluSerIysTyrIysArgSerSerGly 480
DB 1752 CGTCAATTAATACAAATATCATCTATTTAAAGAAAGCAAAATTAACGATCAAGTGT 1811
QY 481 GlyLeuProAlaProSerAspPheIysCysProIleIysGluGluIleAlaIleThrSer 500
DB 1812 GGGCTGCTGCTCCCAAGATGATTTCAAGTGTCTATCAAGAGAGATAGCAATTAACAGT 1871
QY 501 GlyGluTrpGluValLeuGluIysArgHisGlySerAsnIleGlnValAspGluValAlaArg 520
DB 1876 GGTGAATGGAAAGTTTGTGGCGGAGTGAATCTTAATATCAAGTTATGAAAGTCAGAAAG 1931
QY 521 LeuValTyrPheGluGluIleThrIysAspSerProLeuGluHisHisIleuTyrValIleSer 540
DB 1932 CTGGATATATTTGAAGGACCAAAAGACTCCCTTTAAGACATCACTGTACGATGAGT 1991
QY 541 TyrValAsnProGlyGluValIleThrArgLeuThrAspArgGlyTyrSerHisSerCys 560
DB 1992 TACGTAATCTCGAGAGGTGCAAGGCTGACGACCTGAGCTGATCATCATCTTCTGCTGC 2051

```

QY 561 lIeSerGlnHisCysaspPhepHeIIeSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db 2052 ATCAGTCAGACACTGTACTTCTTATAGTAACTAGTAAACAGAAAGATCCACTGT 2111
QY 581 ValSerLeuTyrLysLeuSerSerProGlnAspAspProThrCysLysThrLysGluPhe 600
Db 2112 GATGCCCTTTACAGAGCTATCAGTCCCTGAGAGATGACCAATGCAAAACAAAGGAATT 2171
QY 601 TTPAlaThrIleLeuaspSerAlaGlyProLeuProAspTyrThrProGlnIlePhe 620
Db 2172 TGGGCGACCATTTTGGATTGAGAGGTCCTCTTCGACATATACCTCTCCAGAAATTTTC 2231
QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db 2232 TCTTTTGAAGAGTACTACTGATTTTACATGTATGGGATCTCTCAAGCCTCATGATCTA 2291
QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
Db 2292 CAGCCTGAAAGAAATATCTCTCTGCTGCTCATATATGAGTCCAGGTCGACATG 2351
QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db 2352 GTGATTAATCGGTTTAAAGAGTCAAGTATTTCCGCTGAATACCTTACCTCTAGGT 2411
QY 681 TyrValValValValIleLeuAsnAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
Db 2412 TATGTGTGTGTAGTATGACACACAGGGGATCTCTGACCGAGGCTTAAATTTGAAGGC 2471
QY 701 AlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnIlyr 720
Db 2472 GCCTTTAATATTAATATG----- 2489
QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
Db 2489 ----- 2489
QY 741 GlyIlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db 2490 ----- 2498
QY 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
Db 2499 GCTGGGGCCCGAGTCACTCTGTGAGTCTTCTATGATACAGATACACGAAACGTTATATG 2558
QY 781 GlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGln 800
Db 2559 GATCACTGACCGAATGAACAGGGCTATTACTTAGCATCTGTGGCATGCAAGCAAGAA 2618
QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnVal 820
Db 2619 AAGTTCCCTCTGAAACCAATCGTTTACTGCTCTTACATGGGTTCCTGGATGAGAAATGTC 2678
QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
Db 2679 CATTTTGGACATACACAGTATATATCTAGATTTTATAGAGAGCGCTGAAACCATATGAT 2738
QY 841 LeuGlnIleTyrProGlnGlnArgHisSerIleArgValProGluSerGlyGlnHisTyr 860
Db 2739 TTACAGATCTATCTCAGAGAGACACAGCATTAAGAGTTCTGATCGGAGAACATTAAT 2798
QY 861 GluLeuHisLeuLeuHisTyrLeuGlnGlnLysAsnLeuGlySerArgIleAlaIleLeuLys 880
Db 2799 GAATCGCATCTTTTGCATCTACTTCAAGAAACCTTGATCAGTATGTGCTCTTAA 2858
QY 881 ValIle 882
Db 2859 GTGATA 2864

```

RESULT 11

AX608745

LOCUS AX608745 4685 bp DNA linear PAT 17-FEB-2003

DEFINITION Sequence 22 from Patent WO0231134.

```

ACCESSION AX608745
VERSION AX608745.1 GI:26404308
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
TITLE Novel serine protease genes related to dppiv
JOURNAL Patent: WO 0231134-A 22 18-APR-2002;
Ferring BV (NL)
FEATURES
source Location/Qualifiers
1..4685
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 1430 a 853 c 991 g 1411 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4685
Score: 4385.50 Matches: 834
Percent Simlarity: 94.56% Conservative: 0
Best Local Similarity: 94.56% Mismatches: 1
Query Match: 93.31% Indels: 48
DB: 6 Gaps: 1
US-10-070-464-1 (1-882) x AX608745 (1-4685)
QY 1 MetalAlaAlaMetGluThrGlnGlnLeuGlyValGlnIlePheGluThrAlaAspCys 20
Db 214 ATGGCAGCAGCAATGGAAACAGAAACAGTGGGTGTGAGATTTTGAATCGGACCTGT 273
QY 21 GlnGluAsnIleGluSerGlnAspArgProLysLeuGlnProPheTyrValGlnArgTyr 40
Db 274 GAGAGATATTTGAATCAGAGATCGGCTTAATTTGAGCTTTTATGTTGAGCGGAT 333
QY 41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db 334 TCTGGAGTCAAGCTTAATAAGCTCTGCGCATACCAAAATATCATGGCTACATGATG 393
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db 394 GCTAAGGCACACATGATTTTCATGTTTGTGAAGAGATGATCCAGATGACCTTATCA 453
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGlnAsnArgGlnAsnThrLeuPheTyrSer 100
Db 454 GACGAATCTATTAACCTTGCATGCTGTGAGAAACAGAAATATACACTGTTTATCT 513
QY 101 GlnIleProLysThrIleAsnArgAlaValLeuMetLeuSerTrpLysProLeuLeu 120
Db 514 GAATTTCCCAAACTATCATATGACAGCAATCTTAATGCTCTCTGGAAACCTCTTTTG 573
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGlnGlnLeuLeuArg 140
Db 574 GATCTTTTCAAGCAACATGACTATGAAATGATTTCTGAGAAAGAAAGACTTATTAAG 633
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
Db 634 GAAAGAAAGCAATGGTGAACAGTCGGAATGCTTCTTACGATTTTCACAAAGGAAGTGA 693
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
Db 694 ACATTTCTGTTTCAAGCGGTAGTGAATTTTATCACTTAAAGATGAGGCGCAAGGA 753
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGlnThrSerCysProAsnIleArgMet 200
Db 754 TTACGCAACCAACCTTTAAGGCCCAATCTAGTGAACCTAGTTGCTCCACATACGAGATG 813
QY 201 AspProLysLeuLysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
Db 814 GATCCAAATATATGACCTGCTGATCCAGACTGGAATGCTTTTATACATACAGATATAT 873

```

QY 221 TrrlleserAenllevalThrArgGluGluArgGluThrTyrValHisasnGluLeu 240
 Db 874 TGGATATCTTAACATCGTAACAGAGAAAGAGAACTCACTTATGTGCACAAATGAGCTA 933
 QY 241 AlaasnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGlu 260
 Db 934 GCCAACATGAGAAAGATGCGACAGATGAGTGGAGTGGTACCTTGTCTCCAGAAATA 993
 QY 261 PheAspArgIYrSerGlyTyrTrpTyrCysProAlaGluThrThrProSerGlyVal 280
 Db 994 TTTGATAGATATCTCGGCTATTTGGTGTGTCCAAAGGTGAACAACATCCAGTGTGT 1053
 QY 281 LysIleLeuAlaGlyIleLeuTyrGluGluAsnAspGluSerGluValGlyIleHisVal 300
 Db 1054 AAAATTTCTTAAGATTCTAATATGAAGAAATGATGATGAGTGAATATTATCATGTT 1113
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
 Db 1114 ACATCCCTTATGTTGGAAACAGAGAGGACATTCATTCGGTATCTTAAACAGGTACA 1173
 QY 321 AlaasnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
 Db 1174 GCAATCTTAAGTCACTTTAAGATGTCAGAAATATGATGATGCTGAGAAAGAGATC 1233
 QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
 Db 1234 ATAAATGTCATAGATTAAGGAACTAATTCACCTTTGAGATTCATATTGGAAGAGATTGAA 1293
 QY 361 TyrIleAlaArgAlaGlyTyrTrpThrProGluGlyLysTyrAlaThrSerIleLeuLeuAsp 380
 Db 1294 TATATTCCTCAGAGTGGATGAGATCTCTGAGGAAATATGCTTGTGCTCATCTACATGAT 1353
 QY 381 ArgSerGlnThrArgLeuGlnIleValIleIleSerProGluLeuPheIleProValGlu 400
 Db 1354 CGCTCCAGAGCTCGCTACAGATGATGTGTATCTCACTGATATTTATCCAGTACGAA 1413
 QY 401 AspAspValMetGluArgGluArgLeuIleGluSerValProAspSerValThrProLeu 420
 Db 1414 GAATGATGTATGGAAGGACAGAGCTCATTTGATGATGCTGATTTCTGAGCCCACTA 1473
 QY 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
 Db 1474 ATATCTATGAAAGAAACACAGACATCTGATTAATATTCATGACATCTTTCATGTTTTT 1533
 QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
 Db 1534 CCCCAGAGTCAAGAGAGAAATTTGAGTTATTTTCTGCTGATGCCAAGCAAGCTTC 1593
 QY 461 ArgHisIleuTyrIleTyrIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
 Db 1594 CGTCATTTTATACAAATTTACATCTATTTTAAAGGAAAGCAAAATATAAACGATCCAGTGT 1653
 QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
 Db 1654 GGGCTCCCTGCTCCAGATGATTTCAAGTGTCTTCAAGAGAGAGATAGCAATTTACAGT 1713
 QY 501 GlyGluTrrpGluValLeuGlyArgHisGlySerAsnIleGluValAspGluValArgArg 520
 Db 1714 GGTGAATGGGAAGTTCTTGCGCGCATGATCTAATATCAAGATTGATGAAGCAGAGG 1773
 QY 521 LeuValTyrPheGluGluThrLysAspSerProLeuGluHisIleLeuTyrValIleSer 540
 Db 1774 CTGGTATATTTTGAAGGACCAAAAGCTCCCTTTAGAGGATCACTGTACGTAAGTCAGT 1833
 QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
 Db 1834 TAGGTAAATCTCGAAGAGGAGCAAGGCTGACGACCGTGGCTACTCAATTTCTTCTCG 1893
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGluLysAsnProHisCys 580
 Db 1894 ATCAGTCACTGATGATCTTTCTTATATAGTATATGATACAGAGAAATCCACACTGT 1953

QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
 Db 1954 GTGTCCCTTTCACAGCTATCAAGTCTGTAAGATGACCACTTGCAAAACAAAGGATTT 2013
 QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
 Db 2014 TGGGCAACCATTTTGGATTTCAGT----- 2036
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
 Db 2036 ----- 2036
 QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGluValGlnLeu 660
 Db 2037 -----CCTCAGTGCAGTTG 2051
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
 Db 2052 GTGAATATCGTGTAAAGAGATCAAGATTTCCGCTTGAAATACCTTAGCTCTTAGGT 2111
 QY 681 TyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
 Db 2112 TATGTGGTGTGATGATGACACACAGGAGTCTGTCAACGAGGCTTAAATTTGAAGGC 2171
 QY 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGluValGluGlyLeuGlyTyr 720
 Db 2172 GCCTTAAATATATAAATGGGTCAAAATGAAATACATACATGCGTGAAGAGCTCCAAAT 2231
 QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
 Db 2232 CTAGCTTCTGAAATGATTCATTTGATTTGATGATGATGATGATGATGATGATGATGAT 2291
 QY 741 GlyValTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
 Db 2292 GAGAGATACCTCTCCCTGATGACATTAATGACAGAGTCAAGATATCTCAGGTTGCTAAT 2351
 QY 761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
 Db 2352 GCTGGGGCCCCAGTCACTCTGTGATCTTCTATGATACAGATACAGGAACTGATATGT 2411
 QY 781 GlyHisProAspGluAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
 Db 2412 GGTCACTCTGACCAAGATGAAACAGGCTATTACTTGAATCTGTGGCCATGCAAGCAGAA 2471
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
 Db 2472 AAGTTCCTCTGACCAATCGTTTACGTCTCTTACATGTGTTCTCGATGAGAAATGTC 2531
 QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
 Db 2532 CATTTGGACATACCAAGTATATTACTGAGTTTATTGAGGCTGGAAGGCCATATGAT 2591
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
 Db 2592 TTACAGATCTATTCCTGAGGACACAGACACAGCATAGAGTTCTGAATCGGAGAAACATTAT 2651
 QY 861 GluLeuHisIleLeuHisTyrTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
 Db 2652 GAACTGCATCTTTTGGACTTACCTTCAAGAAACCTTGATGATCATGTGCTGCTTAAAA 2711
 QY 881 ValIle 882
 Db 2712 GTGATA 2717

RESULT 12
 AX608743 4676 bp DNA linear PAT 17-FEB-2003
 LOCUS DEPOSITION
 AX608743 Sequence 20 from Patent WO0231134.
 AX608743
 AX608743.1 GI:26404307
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
Ferring BV (NL)

1
Qi, S., Akinsanya, K.O., Riviere, P.J. and Junien, J.L.
Novel serine protease genes related to dppiv
Patent: WO 0231134-A 20 18-APR-2002;

FEATURES

Source
Location/Qualifiers
1..4676
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT
1424 a 859 c 979 g 1414 t

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4676
Score: 4385.00 Matches: 831
Percent Similarity: 94.22% Conservative: 0
Best Local Similarity: 94.22% Mismatches: 1
Query Match: 93.30% Indels: 51
Gaps: 1

US-10-070-464-1 (1-882) x AX608743 (1-4676)

QY 1 MetlaaiaaamergluThrgluGlunleuglyValgluilepnegluThralaaspCys 20
Db 214 ATGGAGAGCAAGAAAGAAACAGCTGGGTGTGTGAGATATTTGAACCGCGACGT 273
QY 21 GluGluaenilegluSerGlnAspArgProlysleuGluProphetyValgluargTyr 40
Db 274 GAGGAGAAATTTGATATCAGAGATCGGCTTAATTGGAGCCTTTTATGTTGAGCGGTAT 333
QY 41 SerTrpSerGlnleuLysLysleuLeuAlaaspThrArgLysTyrHisglYTYrMetMet 60
Db 334 TCCGTGAGTCAAGCTTAAAGCTGCTTCCGATACCAAGAAATATCATGCTATCATGATG 393
QY 61 AlalyslaaProHisaspPheMetPheValLysArgAspProaspGlyProHisSer 80
Db 394 GCTAAGGACCACTGATTTATGTTGTGAAGAGAAATGATCCAGATGACCTCATTTCA 453
QY 81 AspArgileTyTYrLysleuAlaMetSerGlyLysAsnArgGluAsnThrLeuPheTyrSer 100
Db 454 GACAGAAATCTATACCTTGCCATGCTGCTGTGAGAAACAGAAATATCATGCTTTATTTCT 513
QY 101 GluileProLysThrileAsnArgAlaAlaValleuMetLeuSerTyrLysProleuLeu 120
Db 514 GAAATTCCTCAAACTATCATATGAGACAGACTTTAATGCTCTTGGAAAGCTCTTTTG 573
QY 121 AspleupheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
Db 574 GATCTTTTTCAGGACACTGGAAGCTATGGAATGTATCTCGAAGAAAGAACTATTAAGA 633
QY 141 GluArgLysArgileGlyThrValglYleAlaSerTyrAspTyrHisglngLysGly 160
Db 634 GAAAGAAACGCATTTGGAACAGTGGAAATGCTCTTACGATTTATCCAAAGAAAGTGA 693
QY 161 ThrPheLeupheGlnAlaGlySerGlyLleTyTHisValLysAspGlyGlyProGlnGly 180
Db 694 ACATTTCTGTTCAAGCGGTAGTGAATTTATCACCTTAAAGATGAGGGCCCAAGAGA 753
QY 181 PheThrGlnGlnProleuArgProAsnLeuValgluThrSerCysProAsnileArgMet 200
Db 754 TTTTACGACAACTTTAAGCCCAACTCTAGTGAACCTAGTTTCTCCACATACGGAAT 813
QY 201 AspProLysleuCysProAlaAspProaspTyrileAlaPheileHisSerAsnAspIle 220
Db 814 GATCCAAATTAATGCTGCTGATCCAGACTGATGCTCTTTATACATAGAACATATAT 873
QY 221 ThrileSerAsnileValThrArgGluGluArgArgLeuThrTyrValHisasnGluLeu 240
Db 874 TGGATATCTTAACATCGTAACAGAAAGAAAGAGACTCATTAATGTGCACAATGAGACTA 933

QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValleuGlnGlu 260
Db 934 GCCAACATGGAAGAAAGATGCCAGATCAGCTGAGTGGTACTTTGTTCTCCAGAAAGAA 993
QY 261 PheAspArgLysSerGlyTYrTrpTyrCysProLysAlaGluThrThrProSerGlyGly 280
Db 994 TTTGATAGATATTCCTGCTATTTGGTGTGTGCCAAAGCTGAACAACTCCAGATGCTGT 1053
QY 281 LysileuArgGlyleuTyrgluGluAsnAspGluSerGlyValGluileHisVal 300
Db 1054 AAAATTTCTTAAGAAATTTATATGAAAGAAATATGATGTGAGTGAATTAATTCATGTT 1113
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db 1114 ACATCCCTTATGTTGGAACAAAGAGGCAATTCATTCCTGTTATCTTAACAGCTGACA 1173
QY 321 AlaAsnProLysValThrPheLysMetSerGluileMetileAspAlaGluGlyValGly 340
Db 1174 GCAATCTCTAAAGTCACTTTTAAGATGTCAAGAAATATGATTTGATGCTGAAGAAAGATC 1233
QY 341 IleAspValileAspLysGluLeuileGlnProheGluileuPheGluGlyValGlu 360
Db 1234 ATAGATGTCATAGTAAGAACTATATTCACCTTTTGATTCATATTTGAAGAGATTGAA 1293
QY 361 TyrileAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerileLeuLeuAsp 380
Db 1294 TATATTCAGAGCTGATGATGATCTCTGAGGAAATATGCTGTGCTCATCTCTACAT 1353
QY 381 ArgSerGlnThrArgLysGluGlnileValleuileSerProGluLeuPheileProValGlu 400
Db 1354 CGCTCCAGACTGCTTCAAGATGATGTGTATCTCACCTGAATTAATTAATCCAGTAA 1413
QY 401 AspAspValMetGluArgGlnArgleuileGluSerValProaspSerValThrProleu 420
Db 1414 GATAGTGTATGAAAGGACAGACTATTTAGTCAAGTGCCTGATTCGTAGCGCACTA 1473
QY 421 IleileTyrgluGluThrThrAspIleTyrileAsnileHisAspIlePheHisValPhe 440
Db 1474 ATTATCTATGAAAGAAACAAGACATCTGATTAATATTCATGATCTTTCATGTTT 1533
QY 441 ProGlnSerHisglGluGluGluileGluPheilePheAlaSerGluCysGlyThrGlyPhe 460
Db 1534 CCCCAGAGTCAAGAGAGAAATTAATTAATTTTCCCTGTGAATCCAAAGAGGTTTC 1593
QY 461 ArgHisLeuTyrglyLysileThrSerileLeuLysgluSerTyrLysArgSerSerGly 480
Db 1594 CGTATTTATACAAATTAATCATCTATTTTAAAGAAAGCAAAATTAACATCCAGTGGT 1653
QY 481 GlyLeuProAlaProSerAspPheLysCysProileLysGluGluileAlaileThrSer 500
Db 1654 GGGCTGCTGCTCCAAAGTGAATTCAGTGTCTATCAAGAGAGATAGCAATTAACAGT 1713
QY 501 GluGluThrProGluValleuGlyArgHisglYSerAsnileGlnValaAspGluValaArg 520
Db 1714 GGTGAATGGGAAGTCTTGGCCGCGCATGATCTTAATTAATCAAGTGTATGAAGTCAGAGG 1773
QY 521 LeuValTyrPheGluGluGlyThrLysAspSerProleuGluHisLeuTyrglyValaSer 540
Db 1774 CTGGTATATTTTGAAGCAACCAAGACTCCCTTTAAGATCACTGTAAGTATGAT 1833
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db 1834 TACGTAATCTCTGAGAGGTGAGCAAGGCTGACGATCCGTGCTACTCACTATTCCTGCTG 1893
QY 561 IleSerGlnHisCysAspPhePheileSerLysTyrSerAsnGluLysAsnProHisCys 580
Db 1894 ATCAGTCAGACACTGTGATCTTTATTAAGTATAGTATAGTAACCAAGAAATTCACATGT 1953
QY 581 ValSerleuTyrglyLysleuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db 1954 GTGTCCCTTAAAGCTATCAAGTCTGAAGATGACCAACTTGCAAAACAAAGGAATTT 2013
QY 601 TyrAlaThrileuAspSerAlaGlyProleuProaspTyrThrProGluilePhe 620


```

Db      2014 TGGGCGACCATTTTGGATTCAGCAGCTCTCTTCTGCTATACCTCCCGAAATTTTC
Qy      621 SerPheGluSerThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db      2074 TCTTTTGAAGAGTACTAGTGGATTTACATTGTATGGAGTCTCTACAGACCTCATGATGCTA 2133
Qy      641 GlnProGlyLysLeuTyrProThrValLeuPheLeuTyrGlyGlyProGlnValGlnLeu 660
Db      2134 CAGCTCGGAAAGAAATATCCATCTGTGCTGATATATGATGCTGCG----- 2180
Qy      661 ValAsnAsnArgPheLeuGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db      2180 ----- 2180
Qy      681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGlnGly 700
Db      2180 ----- 2180
Qy      701 AlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnTyr 720
Db      2181 -----GATCAAAATGAAATGACATGACAGTGGAGAGAGCTCCAAATAT 2222
Qy      721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
Db      2223 CTAGCTTCTCGATATGATTCATTGACTTACATGATCGTGGGCATCCACGGCTGATCTAT 2282
Qy      741 GlyGlyTyrLysSerLeuMetAlaLeuMetGlnLysSerAspIlePheArgValAlaIle 760
Db      2283 GAGGAGATCTCTCCCTGATGAGCATTAAGCAGAGGTCAGATATCTTACAGGATTCATAT 2342
Qy      761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlnArgTyrMet 780
Db      2343 GCTGGGGGCCCGCATCTCTGATCTTCTATGATCAGAGTACACAGGAAGTTATATG 2402
Qy      781 GlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGln 800
Db      2403 GGTCAACCTCGACCAAGATGAACAGAGGCTATTAATTAGATCTGTGGCATCAAGCAGAA 2462
Qy      801 LysPheProSerGlnProAsnArgLeuLeuLeuHisGlyPheLeuAspGlnAsnVal 820
Db      2463 AAGTTCCTCTCGAACCAATGCTTACTCTTCAATGCTTCTTCTGATGAGAAATGTC 2522
Qy      821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
Db      2523 CATTTTGGACATACCAAGATATATTACTGATTTTATGAGAGGCTGGAAGCCATATGAT 2582
Qy      841 LeuGlnIleTyrProGlnGlnArgHisSerIleArgValProGlnSerGlyGlnHisTyr 860
Db      2583 TTACAGATCTATCTCGAGAGACACAGCATTAAGATTCCTGATTCGGAGAACATATAT 2642
Qy      861 GlnLeuHisLeuLeuHisTyrLeuGlnGlnAsnLeuGlySerArgIleAlaAlaLeuLys 880
Db      2643 GAATCGCATCTTTTGCATCTTCAAGAAACCTTGATCCAGTATTCGCTGCTTAATAA 2702
Qy      881 ValIle 882
Db      2703 GTGATA 2708

```

RESULT 13
 AX405770 2842 bp DNA linear PAT 14-JUN-2002
 LOCUS Sequence 185 from Patent WO0222660.
 ACCESSION AX405770
 VERSION AX405770.1 GI:21438979
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 1 Tang, Y. T., Liu, C., Zhou, P., Asupdi, V., Zhang, J., Zhao, Q. A., Ren, F.,
 Xue, A. J., Yang, Y., Wehrman, T. and Drmanac, R. T.

```

TITLE      Novel nucleic acids and polypeptides
JOURNAL    Patent: WO 0222660-A 185 21-MAR-2002;
           HYSBO, INC. (US)
FEATURES
  source
    Location/Qualifiers
      1..2842
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        234..2582
          /note="unnamed protein product"
  CDS
    1
      /protein_id="CAD34877.1"
      /db_xref="GI:21438980"
      /translation="MAAMETBQLGVEIFETADCEENIESQDRPKLEPVVRYMSO
      LKULADIRKTHGYMAKAPHDPMFVKRNDPPGPHSDRIITYLAMSCEENENTLFSSEI
      PKTINRAVALMLSWKPLDLFQATLDYGVASBEELREKKIGVGLASVDHQSOG
      TELFQAGSGIYHVKDGPQFTQOPRLVITSCNIMIDPLCPADPDMAFIHSN
      DIMSNIVTREERRLTYVNEELANMEDSRASGATFVLOEERDRSGYWCPEAKETT
      PSQGLIRLIVRENDSEVEIIVHTSPMLSTRADSFYKMTGTAMPKYTPKSEIMI
      DAGRIIDVIDKELIOPFELFPGVEYVIRAGMTPEKXAMSLILDSOTRQIVLIS
      PELFIPVEDDVNERQRLISVPSVTPILIIYETTDIYINIDIFPVPSQHEEIEP
      IFASBCKTFRHLVYKTSILKESKRSGLPAPSDFCPIKEEATISGEVEIAR
      HGSNIQVDEVRRLVYEGTKDSPLEHLLVVSIVNGETRLTDROYSHSCISQCHD
      FPIKSYSNQKHPICVSLYKLSSEDDPTCKTFEVALTILDSAGPLDPDTPPIFSPE
      TTGFTLYGMLYKPHDLPQPKKYPTVLFYGGPQVALAGAPVTLIMFYDTGYERVWGH
      PDNREGYIYGVAMQAEKPEPSPNLIILHGLDENVHFAHTSILSFLVRAKPEYD
      LQYPERHSIVPSGHEHYELHLHYLGENISRTALVKVI"
BASE COUNT      857 a      592 c      635 g      758 t
ORIGIN
Alignment Scores:
  Pred. No.:      0      Length:      2842
  Score:      4118.00      Matches:      782
  Percent Similarity:      88.66%      Conservative:      0
  Best Local Similarity:      88.66%      Mismatches:      0
  Query Match:      87.62%      Indels:      100
  DB:      6      Gaps:      1
US-10-070-464-1 (1-882) x AX405770 (1-2842)
Qy      1 MetAlaAlaLeuMetGlnThrGlnGlnGlnGlyValGlnIlePheGlnThrAlaAspCys 20
Db      234 ATGGCAGCAGCAATGGAAGAACAGCTGGGGTGGAGATATTGAACTGGGACCTG 293
Qy      21 GlnGlnAsnIleGlnSerGlnAspArgProLysLeuGlnProPheTyrValGlnArgTyr 40
Db      294 GAGGAGATATTTGAATCAACAGATCCGCCCTTAATTGAGCTTTTATGTTGACCGGAT 353
Qy      48 SerTyrSerGlnLeuLysLysLeuLeuAlaSerThrArgLysTyrHisGlyTyrMetMet 60
Db      354 TCCCTGAGTACGCTTAAAGCTGCTTCGGAATCAAGAAATATCATGTGCTCAATGATG 413
Qy      61 AlAlaValProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db      414 GCTAAGGACCAACATGATTTTCAATGTTTGTGAGAGAGAAATGATCAATGGAACCTCAT 473
Qy      81 AspArgIleTyrTyrLeuAlaMetSerGlyGlnAsnArgGlnAsnThrLeuPheTyrSer 100
Db      474 GACAGATCTATTAACCTTGCATGCTGGGAGAAACAGAAATATCACTGTTTATTTCT 533
Qy      101 GlnIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeuLeu 120
Db      534 GAAATTCCAAACTTCAATTAAGACAGCGATTTATGCTCTCTGAAAGCCTCTTTTG 593
Qy      121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGlnGlnGlnLeuLeuArg 140
Db      594 GATCTTTTTCAGGCAACACTGACCTATGGAATGATTTATTCGAGAGAAAGAACTATTAGA 653
Qy      141 GlnArgGlnArgGlnIleGlnThrValGlyIleAlaSerTyrAspTyrHisGlnGlnSerGly 160
Db      654 GAAAGAAAGACATTTGAAACAGTCGGAATTCCTTTCAATGATTAATCAACAGAGAAATGGA 713
Qy      161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValIleAspArgIleGlyProGlnGly 180

```

Db	714	ACATTTCTGTTTAAAGCCGGTAGTGAATTTATACGTAAAGATGAGGCGCACAGAA	773		Db	1794	CTGGATATATTTGAAAGCACCAAGACTCCCTTTAGACATACCTGTACGTACAGT	1853
Qy	181	PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet	200		Qy	541	TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys	560
Db	774	TTTACCAACAACCTTTAAGGCCCATCTAGTGGAAACTAGTGTCTCCACATACCGAAG	833		Db	1854	TACGTAAATCTCGGAAGGGTGAACAAGGCTGACGACCGGCTACTACATCTTGGCTGC	1913
Qy	201	AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle	220		Qy	561	IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys	580
Db	834	GATCCAAAATTAGCCCTGCTGATCCAGACTGGATTGCTTTATACATGACCAACGATATT	893		Db	1914	ATCAGTCAGCAGCTGACCTCTTTTATAGTAAAGTATAGTAACCAAGAAATCCACACTGT	1973
Qy	221	TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu	240		Qy	581	ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe	600
Db	894	TGGATTCTTAACATCGTAACCAAGAAAGAAAGAGACTCACTTATGTGACATAGCTCA	953		Db	1974	GTGTCCCTTTAACAAGCTATCAAGTCTCTGAAGATGACCCAACTTGCAAAACAAGAAATT	2033
Qy	241	AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu	260		Qy	601	TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe	620
Db	954	GCCACATGGAAGAAAGATCCAGATCAGCTGGAGTGGCTACCTTGTCTTCCAAAGAA	1013		Db	2034	TGGGCCACCATTTGGATTCAACAGAGTCTCTTCCGATATACCTCCACAGAAATTTTC	2093
Qy	261	PheAspArgTyrSerGlyTyrTyrTrpCysProLysAlaGluThrThrProSerGlyGly	280		Qy	621	SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu	640
Db	1014	TTTGATAGATATTCTGGCTATTGGTGTGTCCAAAAGCTGAAACAACACTCCAGTGGTGT	1073		Db	2094	TCTTTTGAAGTACTACGTGATTTACATTTGATGGGAATGCTCTACAAAGCTCATGATCTA	2153
Qy	281	LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal	300		Qy	641	GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu	660
Db	1074	AAAAATTTCTAGAAATCTATATAGAAAGAAATGATGAAATCGAGTGGAAATTTATCATGTT	1133		Db	2154	CAGCTGGAAGAAATATCTTACTGTGCTGTTCATATATGCTGCTCAG-----	2204
Qy	301	ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr	320		Qy	661	ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly	680
Db	1134	ACATCCCTTATGTGGAAACAAGAGGCGAGATTCACTCCGTTATCTTAAACAGGTACA	1193		Db	2204	-----	2204
Qy	321	AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle	340		Qy	681	TyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly	700
Db	1194	GGAATTCCTAAAGTCACTTTTAAAGTCAAGAAATGATGAAATTCGTCGAAGAAAGCATC	1253		Db	2204	-----	2204
Qy	341	IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlnGlyValGlu	360		Qy	701	AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr	720
Db	1254	ATAGATGTCATAGTAAGAACTAATTCACCTTTTGAGATTCTAATTTGAAGAGATGGA	1313		Db	2204	-----	2204
Qy	361	TyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrValaTyrSerIleLeuLeuAsp	380		Qy	721	LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr	740
Db	1314	TATATTGCCAGAGCTGATGAGACTCCTGAGGGAATATGCTGTGTCATCTTACTAGAT	1373		Db	2204	-----	2204
Qy	381	ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu	400		Qy	741	GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle	760
Db	1374	CGCTCCAGACTCGCTACAGTACTGTTGATCTCACTGAAATTTTATATCCAGTAGAA	1433		Db	2205	-----GTTCTATT	2213
Qy	401	AspAspValMetGluArgGluArgLeuIleGluSerValProAspSerValThrProLeu	420		Qy	761	AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet	780
Db	1434	GATGATGTTATGGAAAGCGACAGACTCATTTGAGTCAAGTCCGTGATTCGTGACGCCACTA	1493		Db	2214	GCTGGGGCCCAAGTCACTGTGGATCTTATGATACAGATACACGGAACGTTATATG	2273
Qy	421	IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe	440		Qy	781	GlyHisProAspGlnAsnGluGlnGlyTyrTyrTyrLeuGlySerValAlaMetGlnAlaGlu	800
Db	1494	ATTATCTATGAAGAAACAACAGACATCTGATTAATATCCATGACATCTTTCATGTTTTT	1553		Db	2274	GGTACCTTGACCAAAATGAAACAAGGCTATTACTTAGAGATCTGGCCATGCCAAGCAGAA	2333
Qy	441	ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe	460		Qy	801	LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnVal	820
Db	1554	CCCCAAATCTCAGAAAGAGAAATGAGTTATTTTGGCTCTGAAATGCAAAACGAGTTTC	1613		Db	2334	AAATTCCCTCTGAACCAAAATCGTTTACTGTCTTACATGATGTTTTCTTGAGAGAAATGTC	2393
Qy	461	ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly	480		Qy	821	HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp	840
Db	1614	CGTCATTATACAAAATTACATCTAATTTAAGAAACAAATTAACGATCCAGTGGGT	1673		Db	2394	CATTTTCACATACAGTAATATTACTGAGTTTTTATGTAGAGGCTGGAAACCATATGAT	2453
Qy	481	GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer	500		Qy	841	LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr	860
Db	1674	GGGTGCTGCTCTCAAGATGATTCAGTGTCTTCAAAAGAGAGATGCAATTACAGT	1733		Db	2454	TTACAGATCTAATCTCTGAGAGACACACAGATTAAGAGTTCTTGAAATGGGGAACATTAAT	2513
Qy	501	GlyGlnTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg	520		Qy	861	GluLeuHisLeuLeuHisIstYrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys	880
Db	1734	GGTAAAGGGAAGTCTTGGCCGGCATGAGATCTAATATCCAAAGTTGATGAGTAGAGG	1793		Db	2514	GAACTGCATCTTTTGACCTACCTTCAAGAAACCTTGGATCAAGATATTGCTGCTTAAA	2573
Qy	521	LeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisHisLeuTyrValValSer	540		Qy	881	ValIle 882	
					Db	2574	GTGATA 2579	

RESULT 14
 AK068731 4523 bp DNA linear PAT 17-FEB-2003
 LOCUS Sequence 8 from Patent WO0231134.
 DEFINITION AX608731
 ACCESSION AX608731
 VERSION AX608731.1 GI:26404301
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 1 Qi, S., Aktsanyva, K.O., Riviere, P.J. and Junien, J.L.
 Novel serine protease genes related to dprlv
 Patent: WO 0231134-A 8 18-APR-2002;
 Ferring BV (NL)
 FEATURES
 source Location/Qualifiers
 1..4523
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 1384 a 828 c 940 g 1371 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 4523
 Score: 4092.50 Matches: 780
 Percent Similarity: 88.44% Conservative: 0
 Best Local Similarity: 87.07% Mismatches: 1
 Query Match: Indels: 102
 DB: Gaps: 1
 US-10-070-464-1 (1-882) x AK608731 (1-4523)
 QY 1 MetAAlAlAlMeCgluThrGluGlnLeuGluValGluIlePheGluThrAlaAspCys 20
 Db 214 ATGGCAGCAGCATATGAAACAGAAACAGCGGATGTGAGATATTTGAAACCTGGCAGCTG 273
 QY 21 GluGluAsnIleGluSerGlnAspArgProGlyLeuGluProPheTyValGluArgTyr 40
 Db 274 GAGGAGAAATTTGAATCAAGATGGCGCTTAATTGAGCCCTTTTATGTTGAGCGGAT 333
 QY 41 SerTyrSerGlnLeuLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetCys 60
 Db 334 TCCTGGAGTCAGCTTAAAGCTGCTTCCGATACCAAGAAATATCATGGCTACATGATG 393
 QY 61 AlAlaAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
 Db 394 GCTAAGGACACATATATTCATGTTTGAGAGAGATGATCCAGATCGACCTCATTC 453
 QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
 Db 454 GACAGATCTTATACCTTCCATGCTGCTGGTGAACAGAGAAATACACTGTTTATTC 513
 QY 101 GluIleProGlyThrLeuAsnArgAlaAlaValIleMetLeuSerTyrLysProLeu 120
 Db 514 GAATTCCTCCAAACATTCATATAGCAGAGCTTATATCTCTCTTGAAGCCCTTTTG 573
 QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
 Db 574 GATCTTTTCAGGCAACATGGACATATGATGATTCGAGAGAAAGAACTATTAGA 633
 QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
 Db 634 GAAAGAAACGATTTGAAACAGTCGGAATTCCTTCTTACGATATTCACCAAGAAAGTGA 693
 QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
 Db 694 ACATTTCTCTTCAAGCCGATGTGAATTTATCACTGATAAAGATGAGAGGCGCAAGAA 753
 QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200

Db 754 TTTACGCAACACCTTTAAGGCCCAATCTAGTGAAGAACTAGTTCACCAACATACGATG 813
 QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
 Db 814 GATCCAAATAATATGATCCCTGCTGATCCAGCTGATTCGATTTATATCATGCAACGATAT 873
 QY 221 ThrIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
 Db 874 TGGATATCTAACATTCGTACCAAGAGAGAGAGACTCATATATGTCACATAGCTA 933
 QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
 Db 934 GCCAATGAGAAAGAAATGCGACATCAGCTGAGCTGCTACCTTTGTTCTCCAGAAAGAA 993
 QY 261 PheAspArgTyrSerGlyTyrTyrProCysProLysAlaGluThrThrProSerGlyGly 280
 Db 994 TTTGATATGATATTCCTGCTATGATGTGTGTCGAAAGCTGAAACACTCCCACTGGTG 1053
 QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleHisVal 300
 Db 1054 AAAATTTCTGAATTTCTATATGAAGAAAGATGAAATCTGAGTGGAATATTATTCATGTT 1113
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
 Db 1114 ACATCCCTATGTTGGAACAGAGAGGCGCATTCATTCCTTATCTTAAACAGGTACA 1173
 QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluArgGly 340
 Db 1174 GCAATTCCTTAAGTCACTTTTAAGATGTCAGAAATATATGATGTCGAGAGAGATC 1233
 QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
 Db 1234 ATAGATGTCATATATGAAGAACTAATTCACCTTTGACATTCATATTTGAAGAGATGGA 1293
 QY 361 TyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrValThrSerIleLeuLeuAsp 380
 Db 1294 TATATTCACAGAGTGTATGATGACTCTGAGAGAAATATCTGTGCTCATCTCATAGAT 1353
 QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
 Db 1354 CGCTCCAGACTGGCTACAGATATGTTGATCTCACTGAATATTATCCACAGTAGAA 1413
 QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
 Db 1414 GATGATGTTTATGAAAGCAGAGACTCATTCATGATGCTGATCTTGTGAGCGCACTA 1473
 QY 421 IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
 Db 1474 ATTATCTATGAAGAAACACAGACATCTGTGAATATTCATGACATCTTTCATGTTTT 1533
 QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
 Db 1534 CCCCAGAGTACAGAGAGAGAAATGATTTATTTTGGCTGTAATGCAGAAACAGGTTTC 1593
 QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
 Db 1594 CGTATTTATACAAATATTCATCTATTTTAAAGAAACAAATATTAACATCCAGTGT 1653
 QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
 Db 1654 GGGCTGCTCCTCCAAATGATTCAGAGTGCCTATCAAAAGAGAGATGCAATTAACAGT 1713
 QY 501 GlyIleThrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
 Db 1714 GGTAAATGGAAGTTCTTGGCCGCGCATGATCTTAATATCCAAAGTATGAAAGTCAAGAG 1773
 QY 521 LeuValTyrPheGluGluThrLysAspSerProLeuGluHisHisLeuTyrValLysSer 540
 Db 1774 CTGCTATATTTTGAAGCAACAAGACTCCCTTTGAAGATCACTGTATCGTATGACT 1833
 QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
 Db 1834 TACGTAAATCTGAGAGGTGACAGAGCTGACCGGTGCTATCACTTCTGTGCTGC 1893

```

QY 561 11eserginhiScyAspPhehej1eserlyTySerasnGlnLysAsnProhiScyS 580
Db 1894 ATCACTGACACTGACTTCTTTTAAAGTAACTAAGTAAACAGAAAGATCCACACT 1953
QY 561 ValSerLeuTyTyLeuSerSerProGluAspAspProthCysLeuThyLysGluPhe 600
Db 1954 GTGTCCTTTTCAAGCTATCAAGTCTCTGAAAGTACCAACCTTGCAAAACAAAGGAAATT 2013
QY 601 TrpAlaThrIleLeuAspSer1aGlyProLeuProAspTyTyThProProGluIlePhe 620
Db 2014 TGGGCAACATTTTGGATTTCAGACAGCTCTCTCTGACCTATCTCTCCGAAATTTTC 2073
QY 621 SerPheGluSerThrThrGlyPheThrLeuTyTyGlyMetLeuTyTyLysProhiSasPhe 640
Db 2074 TCTTTTGAAGACTACTGACTGATTTTACATTTGATGGAGTCTCTCAACAGCTCATATCTA 2133
QY 641 GlnProGlyLysLysTyTyProThrThrAlaLeuPheIleTyTyGlyGlyProGlnValGlnLeu 660
Db 2134 CAGCTGGAAAGAAATATCTTACTGTGTCTTGCATATATGGTGT----- 2178
QY 661 ValAsnAsnArgPheLysGlyValLysTyTyPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db 2178 ----- 2178
QY 681 TyrValValValValIleAspAsnArgLysCysHisArgGlyLeuLysPheGluGly 700
Db 2178 ----- 2178
QY 701 AlaPheLysTyTyLysMetGlyGlnIleGluIleAspAsnGlnValGlnGlyLeuGlnIle 720
Db 2178 ----- 2178
QY 721 LeuAlaSerArgTyTyAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTy 740
Db 2178 ----- 2178
QY 741 GlyGlyTyTyLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db 2179 ----- CGGTT-GCTATT 2189
QY 761 AlaGlyAlaProValThrLeuTyTyIlePheTyTyAspThrGlyTyTyThGlnArgTyMet 780
Db 2190 GCTGGGGGCCAGCACTCTGTGATCTTCTATATACAGATACAGGAAAGCTTATATG 2249
QY 781 GlyHisProAspGlnAsnGlnGlnGlyTyTyTyLeuGlySerValAlaMetGlnAlaGlu 800
Db 2250 GGTCAACCTGACAGAAATGAAACAGGCTATTAGATCTGTGGCATTGCAAGCAAGA 2309
QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db 2310 AAGTTCCTCTGAAACCAAAATGTTTACTGCTTACATGGTTTCTCGATGAGAAATGTC 2369
QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyAsp 840
Db 2370 CATTTGCAATACAGATATATATCTGAGTTTTTTAAAGTGAAGGCTGGAAGCAATATGAT 2429
QY 841 LeuGlnIleTyTyProGlnGlnArgHisSerIleArgValProGluSerGlyGlnHisTy 860
Db 2430 TTACAGATCTATCTCTCAGAGAGACACAGCATAGAATCTGATGGGAGAACATAT 2489
QY 861 GluLeuHisLeuLeuHisTyTyLeuGlnGlnLeuLysLeuArgIleAlaAlaLeuLys 880
Db 2490 GAACTGATCTTTTGCACACTTCAAGTCAAGAAACCTTGATCAGCTATGTGCTCTTAAA 2549
QY 881 ValIle 882
Db 2550 GTGATA 2555

```

```

RESULT 15
AK000290
LOCUS Homo sapiens cDNA FLJ20283 fis, clone HEP04088.
DEFINITION AK000290 2778 bp mRNA linear PRI 22-FEB-2000

```

```

ACCESSION AK000290
VERSION GI:7020272
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Oka,T.,
Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2778)
Sugano,S., Suzuki,Y., Oka,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction: 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1..2778
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEP04088"
/cell_line="HePG2"
/cell_type="hepatoma"
/clone_id="HEP"
/notes="cloning vector pME18SFJ3"
163..1572
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAA91059.1"
/db_xref="GI:7020273"
/translation="MAAMETPOLGVHIFETADCHENIESODRKLPEPFYVERSKQ
LKULADPTKRCHGYMAKAPHDMPFRKNDPDGHSRIITYLMSGENRENTLYSEI
PKTINRAVLMLSKPLDLFOATLDGYMSREBELREKRIGTVAIASDYQSGG
TFLEQASGLYHWKDGPOGFTQPLDPNIVTSCPIRMDPKLCPADPMIAFIHSH
DIVISNIVTEREERLTYVHNELAMMEDASAGVATVLOEFPKRGYGYWCPAEFT
PSGKILRIYENDESEVEIHYTSPMLSTRADSRVRYKGTFRANKVTPKSEIMI
DAGRIIDVIDKELIOPFEILFEGVYIAGVTPKCKVMSIILDSQRLQIVLIS
PELFIVEDVMKRLIIESVPSPVPLIYEETTDIWINIHDFHVPQSHETIEF
IRPSECTGRHRLIKYITST"
BASE COUNT 856 a 562 c 615 g 745 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2778
Score: 4091.00 Matches: 781
Percent Similarity: 88.55% Conservative: 0
Best Local Similarity: 88.55% Mismatches: 1
Query Match: 87.04% Indels: 101
DB: 9 Gaps: 1
US-10-070-464-1 (1-882) x AK000290 (1-2778)
QY 1 MetAlaAlaLeuGlnThrGlnGlnLeuGlyValGlnIlePheGlnThrAlaAspCys 20
Db 163 ATGGCAGCAGCATGGAACAGAAACAGCTGGGTGTTGAATATTGAACTGGGACTCT 222
QY 21 GlnGluAsnIleGluSerGlnAspArgProLysLeuGlnProPheTyTyValGluArgTy 40
Db 223 GAGGAGAAATATGGAATCAGAGATCGGCTTAATGAGCTTTTATGTGAGCGGTAT 282

```

```

QY 41 SerTSPserGlnLeuLysLysLeuAlaaspThrArgLysTyrHisGlyTyrMetMet 60
Db 283 TCCTGGAGTCAGCTTAAAAAGCTGCTGCATACAGAAAATATCATGCTACATGANTG 342
QY 61 AlAlValAlaProHisAspPheMetPheValLysArgAspProAspGlyProHisSer 80
Db 343 GGTAAAGGACCAAGATTCATGTTTGGAAGAGAAATGATCCAGATGAGCTCATCA 402
QY 81 AsparGlyLeuTyrTyrLeuAlaMetSerGlyLysAsnArgGluAsnThrPhePheSer 100
Db 403 GACGAACTCTATTACTTCCCTGCGCATGCTGCTGGAGAACAGAAAATACACTGTTTATCT 462
QY 101 GluLysProLysThrLysAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeuLeu 120
Db 463 GAAATTCCTCAAAACATCATATAGACACAGCTTAATGCTCTTGGAAAGCTCTTTTG 522
QY 121 AspheuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
Db 523 GATCTTTTCAGGCAACCTGACTATGAAATGTAATTCCTCGAAGAAAGAACTTTAAGA 582
QY 141 GluArgLysArgLysGlyThrValGlyLeuAlaSerTyrAspTyrHisGlyGlySerGly 160
Db 583 GAAAGAAAAGCATTTGGAACAGCTGCAATTGCTTTACGATTATCACCAAGAAAGTGA 642
QY 161 ThrPheLeuPheGlnAlaGlySerGlyLysTyrHisValLysAspGlyGlyProGlnGly 180
Db 643 ACATTTCTGTTTCAAGCCGTAGTGAATTATCATCGTAAAGATGAGGCGCCACAAGGA 702
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnLysArgMet 200
Db 703 TTTACGACAAACCTTTTAGGCCCAATTTAGTGAAGACTAGTGTCCCAACATCGAGATG 762
QY 201 AspProLysLeuCysProAlaAspProAspTyrLysAlaPheLeuHisSerAsnAspLys 220
Db 763 GATCCAAAATTATGCTCCCTGCTGATCCAGATCGATGCTTTTATACATAGAACATTT 822
QY 221 TyrLysSerLysLeuValThrArgGluGluArgLysLeuTyrValHisAsnGluLeu 240
Db 823 TGGATATCTTAACATCGTAAACAGAAAGAAAGAGACTCATTTATGTCACAAATGACTA 882
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
Db 883 GCCAACTGGAAGAAAGATGCCAAGTCCAGCTGAGAGTCCCTGTTGTTCCCAAGAAAGAA 942
QY 261 PheAspArgTyrSerGlyTyrTyrTyrCysProLysAlaGluThrThrProSerGlyGly 280
Db 943 TTTGATAGATATCTGCTATGCTATGCTGATGCCAAAGCTGAACAACTCCCAAGTGGT 1002
QY 281 LysGluLeuArgLysLeuTyrGluGluAsnAspGluSerGluValGluLysLeuHisVal 300
Db 1003 AAAATTTCTTGAATTTCTATATGAAAGAAATGATGAACTGAGTGGAAATTTATCATGTT 1062
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db 1063 ACATCCCTCTATGTTGGAACAAAGAGGCGCATTCATTCGCTATCTCTAAACAGGTACA 1122
QY 321 AlaAsnProLysValThrPheLysMetSerGluLysMetLysAspAlaGluLysArgLys 340
Db 1123 GCAAAATCTTAAAGTCACTTTTAAGATCTCAGAAATATATGATGCTGGAAGAGGATC 1182
QY 341 LysAspValLysLysLysGluLeuLysGlnProPheGluLysLeuPheGluLysValGlu 360
Db 1183 ATTAGATGCTATATAGAAATGAAATTCATTCACCTTTGAGATTCATTTTGAAGAGATTGAA 1242
QY 361 TyrLysAlaArgAlaGlyTyrThrProGluGlyLysTyrValAlaThrSerLysLeuLeuAsp 380
Db 1243 TATATGCCAGAGCTGAGTGAATCTCTGAGGGAATATAGCTTGCTCATCTACATAGAT 1302
QY 381 ArgSerGlnThrArgLysGlnLysLeuLysSerProGluLeuPheLysProValGlu 400
Db 1303 CCTCCCAAGACTGCTACAGATAGATGCTGATCTCACTGAATTTATTTACCAAGTAGAA 1362

```

```

QY 401 AspaSpValMetGluArgGlnArgLysLeuLysSerValProAspSerValThrProLeu 420
Db 1363 GATGATGTTATAGGAAGACAGAGACTCATTTGATGATAGAGCTGATTTCTGTAGCGCCACTA 1422
QY 421 IleLysTyrGluGluThrThrAspLysLeuLysLeuHisAspLysPheHisValPhe 440
Db 1423 ATTATCTATGAAAGAAACAACAGACATCTGATTAATATCCATGACATCTTTCATGTTTTT 1482
QY 441 ProGlnSerHisGluGluGluLysLeuPheLysPheAlaSerGluCysLysThrGlyPhe 460
Db 1483 CCCCAGATGTCAGAGAGAGAAATGAGTTATTTTGGCTCTGAAATGCAAAACAGGTTTC 1542
QY 461 ArgHisLeuTyrLysLysLeuThrSerLysLeuLysGluSerLysTyrLysArgSerGly 480
Db 1543 CGTCAATTATACAAATATACATCTAT-TTAAAGGAAACCAATATATAAAGATCCAGTGGT 1601
QY 481 GlyLeuProAlaProSerAspPheLysCysProLysGluGluLysAlaAlaLysThrSer 500
Db 1602 GGGCTGCTGCTCCAGATGATTTCAAGTGTCTCTATCAAGAGAGATAGCAATTTCCAGT 1661
QY 501 GlyGluTyrProGluValLeuGlyArgHisGlySerAsnLysGlnValAspGluValArgArg 520
Db 1662 GGTAAATGGGAAGTCTTGGCCGCGCATGATCTATATCCAAAGTTGATGAAGTCAGAAAG 1721
QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisLysLeuTyrValValSer 540
Db 1722 CTGCTATATTTTGAAGGACCAAGACCTCCCTTTAGAGCATCACCTTATGATGATCTGAGT 1781
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db 1782 TACGTAAATCCTGAGAGAGTGACAAAGCTGATGACCGTGGCTACTCAATTTCTTGCTGC 1841
QY 561 IleSerGlnHisCysAspPhePheLysSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db 1842 ATCAGTACAGACCTGTGACTTCTTTATAGTAAAGTATAGTAAACCAAGAAATCCACACTG 1901
QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db 1902 GTGTCCCTTTTCAAGCTATCAAGTCTGAAAGATGACCCACTTGGAAAAACAAGGAATTT 1961
QY 601 TyrAlaThrLysLeuAspSerAlaGlyProLeuProAspTyrThrProProGluLysPhe 620
Db 1962 TGGGCCACCAATTTGATGATTCAGAGAGTCCCTCTTCTGACTATACTCTCCCAAGAAATTTT 2021
QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db 2022 TCTTTGAAAGTACTACTGATTTATCATTTGATGAGATGCTCTCAAGCCTCATGATCTCA 2081
QY 641 GlnProGlyLysLysTyrProThrValLeuPheLysTyrGlyGlyProGlnValGlnLeu 660
Db 2082 CAGCTCGAAGAAAGAAATATCTCACTGCTGCTGTTCAATATAGTGGTCCCTCAG----- 2132
QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db 2132 ----- 2132
QY 681 TyrValValValValLysAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
Db 2132 ----- 2132
QY 701 AlaPheLysTyrLysMetGlyGlnLysLeuLysAspGlnValGluGlyLeuGlnTyr 720
Db 2132 ----- 2132
QY 721 LeuAlaSerArgTyrAspPheLysAspLeuAspArgValGlyLysHisGlyTyrSerTyr 740
Db 2132 ----- 2132
QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspLysPheArgValAlaLys 760
Db 2133 -----GTTGGCTATT 2141
QY 761 AlaGlyAlaProValThrLeuTyrLysPheTyrAspThrGlyTyrThrGluArgTyrMet 780

```

```

Db      2142 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGATACACGGAACCTTATATG 2201
Qy      781 G|YH|SP|O|A|S|P|G|I|n|a|n|g|I|n|g|I|Y|T|Y|T|Y|L|e|u|G|I|S|e|V|a|I|a|n|e|G|I|n|a|I|u 800
Db      2202 GGTCAACCCCTGACACAGATGAAACAGGCTATTACTTAGATCTGTGCCATGCAAGCAGAA 2261
Qy      801 LysPheProSerGlnProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db      2262 AAGTTCCTCCTGGAACCAATCGTTTACTGCTCTTACATGATGTTCTCTGATGAGAAATGTC 2321
Qy      821 H|S|P|H|e|A|H|e|T|H|S|e|I|l|e|u|L|e|u|S|e|r|P|H|e|L|e|u|V|a|I|A|G|I|Y|L|S|P|P|o|T|Y|r|A|S|P 840
Db      2322 CATTTTGGACATACCAAGATATATTACTGAGTGTTTTAAAGTGAGGCTGGAAGGCCATATGAT 2381
Qy      841 L|e|u|G|I|n|I|e|T|Y|r|P|r|o|G|I|n|I|u|A|R|G|H|S|e|r|I|e|A|R|V|a|I|P|r|o|G|I|S|e|r|G|I|u|H|I|S|T|Y|r 860
Db      2382 TTAACAGATCTATCTCAGGAGAGACACAGCATTAAGSTTCCTGAATCGGAGAAACATTAT 2441
Qy      861 G|I|L|e|u|H|I|S|L|e|u|H|I|S|T|Y|r|L|e|u|G|I|n|I|u|A|S|n|L|e|u|G|I|S|e|r|A|R|G|I|l|e|A|I|a|I|a|L|e|u|L|Y|S 880
Db      2442 GAACTGCATCTTTTGCACCTACCTCAAGMAAACCTTGATCAAGTATGCTGCTCTTAAAA 2501
Qy      881 V|a|I|I|e 882
Db      2502 GTGATA 2507

```

Search completed: October 16, 2003, 00:03:57
 Job time : 9759.41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:14:56 ; Search time 648.054 Seconds

(Without alignments)
3673.932 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700

Sequence: 1 MAAMMETRQGLVEIFETADG.....HLHYLDENLGSRIALKVT 882

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10070464/rumat.15102003.113552.24799/app.query.fasta_1.2652
-DB=N_Geneseq.19Jun03 -OPMT=fastcap -STUFFIX=ring -MINMATCH=0.1 -LOOFCI=0
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=p2n -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US10070464.@CGN.1.1.0.@rnat.15102003.113552.24799 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq.19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4700	100.0	2649	25	ABX12255	cDNA encoding huma
2	4700	100.0	2671	24	ABK83322	cDNA encoding huma
3	4700	100.0	3106	24	ABK12892	Human protease PRT
4	4700	100.0	3120	22	AA085694	Nucleotide sequenc
5	4700	100.0	3120	24	AA085696	Human dipeptidyl p
6	4700	100.0	3143	24	AA099934	cDNA encoding 2195
7	4695	99.9	2643	24	AA099935	Coding sequence of
8	4680	99.6	4829	24	ABK83327	cDNA encoding huma
9	4385.5	93.3	4685	24	ABK83332	cDNA encoding huma
10	4385	93.3	4676	24	ABK83331	cDNA encoding huma
11	4118	87.6	2842	24	ABN59774	Novel human coding
12	4092.5	87.1	4523	24	ABK83325	cDNA encoding huma
13	3970.5	84.5	2510	24	AA023843	Human protease PRT
14	3771	80.2	2668	24	ABN59775	Novel human coding
15	3661.5	77.9	4309	24	ABK83328	cDNA encoding huma
16	3564.5	71.6	2161	22	AAH15009	Human cDNA sequenc
17	2870	61.1	2617	24	ABK83323	Human cDNA sequenc
18	2870	61.1	3716	24	ABO75955	Human cDNA sequenc
19	2870	61.1	4219	24	ABK83335	Human cDNA sequenc
20	2870	61.1	4302	24	ABK83333	Human cDNA sequenc
21	2863	60.9	3024	24	AA085694	Human cDNA sequenc
22	2835	60.3	2495	24	AA085697	Human cDNA sequenc
23	2833	60.3	3287	24	AA085695	Human cDNA sequenc
24	2820.5	60.0	4180	24	ABK83339	Human cDNA sequenc
25	2820.5	60.0	4263	24	ABK83338	Human cDNA sequenc
26	2801	59.6	1821	24	ABV76411	Human cDNA sequenc
27	2763	58.8	2751	24	AA083311	Human cDNA sequenc
28	2649	56.4	4076	24	ABK83337	Human cDNA sequenc
29	2649	56.4	4159	24	ABK83336	Human cDNA sequenc
30	2638	56.1	2801	22	AA157896	Human cDNA sequenc
31	2599.5	55.3	4037	24	ABK83341	Human cDNA sequenc
32	2599.5	55.3	4120	24	ABK83340	Human cDNA sequenc
33	2476.5	52.7	3262	22	AA157880	Human cDNA sequenc
34	2422	51.5	1669	22	AA085696	Human cDNA sequenc
35	2226.5	47.4	2982	24	AA159666	Human cDNA sequenc
36	1914.5	40.7	2461	21	AA085695	Human cDNA sequenc
37	1878.5	40.0	2952	24	ABK69090	Human cDNA sequenc
38	1874.5	39.9	3047	24	ABK69113	Human cDNA sequenc
39	1836.5	39.1	1083	22	AA085697	Human cDNA sequenc
40	1644.5	35.0	1197	22	AA085695	Human cDNA sequenc
41	1644.5	35.0	2027	21	AA085697	Human cDNA sequenc
42	1599.5	34.0	3733	23	ABL10425	Human cDNA sequenc
43	1599.5	34.0	3783	23	ABL10425	Human cDNA sequenc
44	1400	29.8	2079	21	AA037672	Human cDNA sequenc
45	1400	29.8	2079	21	AA037672	Human cDNA sequenc

ALIGNMENTS

RESULT 1

ABX12255 standard; cDNA; 2649 BP.

ABX12255;

19-MAY-2003 (First entry)

cDNA encoding human serine protease HIPHM46.

KW Human; ss; gene; HIPHM46; serine protease; gene therapy; osteoarthritis;
KW serine protease activity; modulation; dipeptidyl peptidase activity;
KW musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
KW Alzheimer's disease; parasympathetic palsy; Huntington's disease;
KW amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
KW irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;
KW haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;

KM colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
 KM multiple sclerosis; chromosome 15q21-q22.

OS Homo sapiens.

XX Key Location/Qualifiers
 XX Key 1..2649
 XX CDS /tag= a
 FT /product= "HIPHM46"
 FT /note= "Serine protease"

XX GB2374869-A.

XX 30-OCT-2002.

XX 22-JAN-2002; 2002GB-0001404.

XX 23-JAN-2001; 2001GB-0001760.

XX (GLAXO) GLAXO GROUP LTD.

XX Edbrooke MR, Lewis AP;

XX WPI; 2003-150703/15.

XX P-PSDB; AB007720.

XX Identifying modulators of serine protease activity useful for treating
 PT muscularkeletal diseases, by contacting cell expressing a novel serine
 PT protease polypeptide with a compound and monitoring serine protease
 PT activity

XX Claim 12; Page 22-26; 38pp; English.

XX The invention relates to a method of identifying a substance that
 CC modulates serine protease activity, comprising contacting a cell such as
 CC a neuronal cell, lung cell, intestinal cell or a cell infected with a
 CC virus, expressing a serine protease polypeptide (HIPHM 46), or its
 CC variant having dipeptidyl peptidase activity, or a serine protease
 CC isolated from the cell with a test substance and monitoring for serine
 CC protease activity. The method is useful for identifying a substance that
 CC modulates serine protease activity. A modulator of the serine protease is
 CC useful in the manufacture of a medicament for treatment or prophylaxis of
 CC a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
 CC infection, Alzheimer's disease, paraspinal nuclear palsy, myotonic
 CC dystrophy, Huntington's disease or amyotrophic lateral sclerosis.
 CC Additional disease that may be treated using modulators of the serine
 CC protease include malabsorption syndromes, irritable bowel syndrome, lung
 CC disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,
 CC rectal polyps, small bowel tumours, colorectal tumours, anaemia,
 CC dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
 CC sclerosis. The present sequence represents cDNA encoding the human serine
 CC protease HIPHM46 which is located on chromosome 15q21-q22.

XX Sequence 2649 BP; 803 A; 514 C; 585 G; 747 T; 0 other;

XX Alignment Scores:

XX Pred. No.: 0 Length: 2649
 XX Score: 4700.00 Matches: 882
 XX Percent Similarity: 100.00% Conservative: 0
 XX Best Local Similarity: 100.00% Mismatches: 0
 XX Query Match: 100.00% Indels: 0
 XX Gaps: 0

XX US-10-070-464-1 (1-882) x ABX12255 (1-2649)

QY 1 MetAlaAlaMetGluThrGluGlnLeuGlyValAluIlePheGluThrAlaAspCys 20
 Db 1 ATGGCAGCGACGATGGAACAGAGACGCTGGGTGTGAGATATTGAACTCGGAGCTGT 60
 QY 21 GluGluAsnIleGluSerGlnAspArgProlGlySerGluProPheTyrValGluArgTyr 40
 Db 61 GAGGAGAAATATGATACACAGAGATCGCTTAATGGACCTTTTATGTGAGCGGTAT 120

QY 41 SerTyrSerGluLeuIleGlySerLeuLeuAlaAspThrArgGlySerTyrHisGlyTyrMetMet 60
 Db 121 TCTCGAGTCAAGCTTAAAGAGCTGCTGCCGATACAGAAATATATCTGGCTACATGATG 180
 QY 61 AlaAlaAlaProHisAspPheMetPheValIleArgAsnAspProAspGlyProHisSer 80
 Db 181 GCTAAGGACCAACATGATTTTCATGTTTGCAAGAGAAATATACAGATGAGACCTCATTTCA 240
 QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
 Db 241 GACAGATCTTATACCTTGCCATGCTGGTGGAGAACAGAAAAATACATGCTTTTATTTCT 300
 QY 101 GluIleProlGlyThrIleAsnArgAlaAlaValLeuMetLeuSerTyrPylProLeuLeu 120
 Db 301 GAAATTCACAAACTATCAATAGAGCAGCTTAATGCTCTTGAGACCTCTTTTG 360
 QY 121 AspleupheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
 Db 361 GATCTTTTTCAGGCACACCTGACTATGATGATGATTTCTCGAAGAAAGAACTTTAAGA 420
 QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlySerGly 160
 Db 421 GAAGAAAGACGACATGGACAGCTGGAATGCTTTACATATTCACAGAGAAAGTGA 480
 QY 161 ThrPheLeuPheGlnAlaGlySerGlyTyrHisValIleAspAspGlyGlyProGlnGly 180
 Db 481 ACATTTCTGTTTCAAGCCGAGTGAATATATACGTAAGAAAGTGAAGGCCACAGAGA 540
 QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
 Db 541 TTTACGCAACACTTTAAGGCCCAATCTAGTGAATCTAGTTGCCACATACGAGATG 600
 QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
 Db 601 GATCCAAATATATTCCTGCTGACACGACGAGTGTGTTTAAACATAGAACAGATAT 660
 QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
 Db 661 TGGATATCTTAACATGACGTAACAGAGAAAGAGACTCACTTATGTCACAAATGACGTA 720
 QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
 Db 721 GCCAATCAGAAAGAAAGATGCCAGATCAGCTGAGTCCGCTTGTCCCAAGAAAGAA 780
 QY 261 PheAspArgTyrSerGlyTyrTyrTrpCysProLysAlaGluThrThrProSerGlyGly 280
 Db 781 TTTGATGATATATTCGCTATTTGGGTGTGCCAAAGCTGAACAACTCCACAGTGTGT 840
 QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValAluIleIleHisVal 300
 Db 841 AAAATCTTAAGAAATCTATGAGAAAGAAATGATGATCTGAGGTGAAATTTATTCATGTT 900
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
 Db 901 ACATCCCTATGTTGGAAACAGAGGCGCAGATTCATCCGTTATCTTAAACAGATGACA 960
 QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
 Db 961 GCAAAATCCTAAAGTCACTTTAAGATGTCAGAAAAAATGATGATGTCGAAAGAAAGATC 1020
 QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
 Db 1021 ATGATGTCATGATGATGAACTAATTTACCTTTTGAGATTCTATTTGAAGAAGATTGA 1080
 QY 361 TyrIleAlaArgAlaGlyTyrThrProGluGlyTyrValAlaTyrSerIleLeuLeuAsp 380
 Db 1081 TATATTCGACAGAGCTGATGATCTCTGAGGAAAAATATGCTGTGCTCATCTACATGAT 1140
 QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
 Db 1141 CGCTCCCAAGCTCGCTCGAGATGATGATGATCTCAATTTTATTTCCAGATGAGA 1200
 QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420

[illegible]

Db	2281	GCTGGGGCCCCAGTCACCTCTGTGGATCTTCTATGATACAGGATACAGGAACGTTATATG	2340		
Qy	781	GlyHisProAspGlnAsnGluGlnGlyTyrTyrIleuGlySerValAlaMetGlnAlaGlu	800		
Db	2341	GGTCAACCTTGACCAAGAAAGAACAGGGCTATTACTTAGAATCTGTGGCCATGCACAGACGAA	2400		
Qy	801	LysPheProSerGluProAsnArgLeuLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal	820		
Db	2401	AAAGTCCCCCTTGAAACCAATGCTTTACTGCTCTTACATGGATTCTTCGAGAGGAATGTC	2460		
Qy	821	HisPheAlaHisIsthSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp	840		
Db	2461	CATTGTGCATACACAGATATATTACTGAGTCTTTTATGTAGGGGCTGGAAAGCCATATGAT	2520		
Qy	841	LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyIuHisTyr	860		
Db	2521	TTACAGACTCTATCTCTCGAGAGAGACACAGCTAAGAAGTCCCTGAATCTGGAGAAACATTAT	2580		
Qy	861	GluLeuHisIsthLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLys	880		
Db	2581	GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCAGTATTGCTGCTTAAAA	2640		
Qy	881	ValIle 882			
Db	2641	GTGATA 2646			
RESULT 2					
ID	ABK83322	ABK83322 standard; cDNA; 2671 BP.			
AC	ABK83322;				
XX	12-AUG-2002	(first entry)			
XX	CDNA	encoding human DPPIV related serine protease DPPP-1.			
XX	Human; Serine	protease; dipeptidyl peptidase IV-related protein; DPPP;			
KW	DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;				
KW	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;				
KW	heart failure; hypertension; urinary retention; osteoporosis; cancer;				
KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;				
KW	dykinesia; reproductive disorder; inflammatory disorder;				
KW	metabolic disorder; gene; ss.				
OS	Homo sapiens.				
XX	WO200231134-A2.				
XX	18-APR-2002.				
XX	12-OCT-2001; 2001WO-US31874.				
XX	12-OCT-2000; 2000US-240117P.				
XX	(FERR) FERRING BV.				
XX	Qi S, Akinsanya KO, Riviere PJ, Junien J;				
XX	WPI; 2002-444178/47.				
XX	P-PSDB; ABG61591.				
New dipeptidyl peptidase IV-related proteins and nucleic acids encoding					
the proteins, useful for treating e.g. fungal, bacterial, protozoan and					
viral infections, cancers, allergies, neurological disorders, or pain					
-					
Claim 1; Page 53-54; 113pp; English.					
The present invention relates to the isolation of novel human serine					
proteases referred to as dipeptidyl peptidase IV (DPPIV)-related					
proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)					
and nucleic acids encoding them are useful for treating infections					
such as fungal, bacterial, protozoan and viral infections, cancers, allergies,					
neurological disorders, or pain					

CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinestias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.

XX Sequence 2671 BP; 805 A; 524 C; 594 G; 748 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	2671
Score:	4700.00	Matches:	882
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-10-070-464-1 (1-882) x ABK83322 (1-2671)

```

QY 1 MetAlaAlaIaIaMetGluThrGluGluValGluIlePheGluThrAlaAspCys 20
DB 8 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGAGATATTGAACTGGCGACTGT 67
QY 21 GluGluAsnIleGluSerGluAspArgProIleGluGluProPheTyrValGluArgTyr 40
DB 68 GAGAGAAATATGATCATCAGATCGGCTTAAATGAGCTTTTATGTTGACCGGTAT 127
QY 41 SerTyrSerGluLeuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 60
DB 128 TCTTGGAGTCACTTAAAGAGCTGCTGCGATACCGAAATATCATGCTCATCATGATG 187
QY 61 AlaAlaAlaProIleAspPheMetPheValIleAspAspProAspGluProIleSer 80
DB 188 GCTAAGGACCAATGATTTTCACTTTGTAAGAGAAATATCAAGTTGCACTTCAATTA 247
QY 81 AspArgIleTyrTyrLeuAlaMetSerGluIleuAsnArgGluAsnThrIleuPheTyrSer 100
DB 248 GACAGATCTATATACCTTGGCATGCTGGTGAGAAACAGAAATATACATGTTTATTTCT 307
QY 101 GluIleProIleThrIleAsnArgAlaIleValIleuMetLeuSerTyrIleProIleu 120
DB 308 GAAATTCCTCAAACTATCAATAGAGAGAGCATCTTAATGCTCTTGGAAAGCCTCTTTTG 367
QY 121 AspIleuPheGluAlaThrIleuAspTyrGluMetTyrSerArgGluGluIleuLeuArg 140
DB 368 GATCTTTTCAAGCAACATGGACCTATGGATGATATCTCGAAGAAAGAACTATTAAAG 427
QY 141 GluArgLysArgIleGluTyrValGlyIleAlaSerTyrAspTyrHisGlnIleuSerGly 160
DB 428 GAAAGAAAACGCAATGGAAACAGTCGAATTCCTTACGATTAATCAACAAAGAAATGGA 487
QY 161 ThrPheLeuPheGluAlaGluSerGlyIleTyrHisValIleuAspGluGluProGlnGly 180
DB 488 AACTTTCTGTTTCAACCGGAGTGGAAATTTATCAAGTAAAGATGGAAGGCAACAGGA 547
QY 181 PheThrGlnGluProIleuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB 548 TTTAGCAACAACTTTAAGCCCAATCTAGTGAACAACTAGTTGTCCTCAACATACGAG 607
QY 201 AspProIleuLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
DB 608 GATCCAAATTTATGCTCTGCTGATCCAGACTGATTCCTTTATACATAGCAACGATATT 667
QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisGlnGlu 240
DB 668 TGGATATCTAATCATGCTAATCAAGAGAAAGAAAGAACTAATATGTCACATAGAGCTA 727
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValIleuGlnGlu 260
DB 728 GCCAACATAGAGAGATGCGAGATCAGTGAAGTCTGACTTGTCTCCAAAGAAAGA 787

```

```

QY 261 PheAspArgTyrSerGlyTyrTrpTyrCysProIleAlaGluThrThrProSerGlyGly 280
DB 788 TTTGATAGATATCTGCTATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 847
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
DB 848 AAAATCTTATGATTTCTATATGAAAGAAATGAAATCTAGGTGGAATTTATTCATGTT 907
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProIleTrpGlyThr 320
DB 908 ACATCCCTATGTTGAAACCAAGAGGCGAGATTCAATTCCTGTTATCCTTAACAACGTA 967
QY 321 AlaAspProIleValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
DB 968 GCAATCTCAAACTCACTTTAAGATGTCAGAAATATATGATGCTGGAAGAGAGATC 1027
QY 341 IleAspValIleAspLysGluLeuIleGluProPheGluIleLeuPheGluGlyValGlu 360
DB 1028 ATGATGTCATATGATAGAACTAATTCACCTTTGAGATTTCTATTGAAAGAGTTGAA 1087
QY 361 TyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
DB 1088 TATATTCAGAGCTGATGATGATCTCTGAGGAAATATGCTTGTCCATCTTACATGAT 1147
QY 381 ArgSerGluThrArgLeuGluIleValIleuIleSerProGluLeuPheIleProValGlu 400
DB 1148 CGCTCCAGACCTGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1207
QY 401 AspAspValMetGluArgIleuIleGluSerValProAspSerValThrProLeu 420
DB 1208 GATGATGTTATGGAAGAGAGAGACCATTTAGTGTGCTGATCTGAGAGCCACTA 1267
QY 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
DB 1268 ATTATCTAAGAAACCAACAGACATCTGATTAATATCATGATCTTTCATGTTT 1327
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
DB 1328 CCCCAAGTCAAGAGAGAGAAATTTGATTTATTTTTCCTGAAAGCAAAAGAGTTTC 1387
QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerTyrLysArgSerGly 480
DB 1388 CGTCAATTTATCAAAATTAATCATCTAATTTAAAGAAAGCAATTAACATCAATGCT 1447
QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
DB 1448 GAGCTGCTGCTGCCAAGTATTTCAAGTCTCTATCAAGAGAGATGCAATTAACAG 1507
QY 501 GlyIleTrpGluValIleuGlyArgHisGluSerAsnIleGluValAspGluValArgArg 520
DB 1508 GGTGAATGGGAAGTCTTGCGCGCATGATCTAATATCAAGTTATGATGAGACAGAG 1567
QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValIleSer 540
DB 1568 CTGGATATTTTGAAGGACCAAGACCTCCCTTAAGACATCACTGATCGTACAT 1627
QY 541 TyrValAspProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
DB 1628 TACGTAATCTCGGAGAGGTGCAAGCTGACGACCGGTGCTACCAATCTTCTGCTCG 1687
QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
DB 1688 ATCACTGACGACATGATCTTCTTTAATGATAGTATAGTAAACAGAGATCAACAGCTGT 1747
QY 581 ValSerLeuTyrLysLeuSerSerProGluAspProThrCysLysThrLysGluPhe 600
DB 1748 GTGTCCCTTTCAAGACTATCAAGTCTGAAGATGACCAACTTGCAAAACAAAGAAATTT 1807
QY 601 TrpAlaTrpIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
DB 1808 TGGGCCACATTTGGATTCAGCAGGTCTCTTCTGACTATATCTTCCACGAAATTTTTC 1867

```

```

OY 621 SerpneGluSerThrThrglyPheThrLeuTyrGlyMetLeuTyrIleProHisAspLeu 640
DB 1868 TCTTTGAAAGTACTCTGATTTACATTGTATGGAGATGCTCTACAAAGCTCAATGATCTA 1927
OY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
DB 1928 CAGCTCGAAGAAATATCTCTACTGCTGCTCATATAGTGTCTCAGGTCGAGCTTG 1987
OY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAsnLeuGly 680
DB 1988 GTGAATTAATCGATTAAAGAGTCAAGTATTTCCGCTTAATACCTTACCTCTAGCT 2047
OY 681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
DB 2048 TATGTGTTGTATGATGATACACACGGGAGTCTGTCACCGAGGCTTAAATTTGAAGGC 2107
OY 701 AlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGlnValGlnLysLeuGlnTyr 720
DB 2108 GCCTTTAAATATTAATATGAGTCAAAATAGAAATTTGACGATCAGTGGAAAGACTCCAAAT 2167
OY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleGlyTyrPheTyr 740
DB 2168 CTAGCTTCTCGATGATGATTTCACTTATGATGATGATGATGATGATGATGATGATGAT 2227
OY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
DB 2228 GGAGGATACCTCTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2287
OY 761 AlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGlnArgTyrMet 780
DB 2288 GCTGGGGCCCGACGACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2347
OY 781 GlyHisProAspGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 800
DB 2348 GGTACCTGTCAGCAATGAAACAGGGCTATTACTTAGATCTGTGGCCATGCAAGCAGAA 2407
OY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 820
DB 2408 AAGTTTCCCTCTGAAACCAATCGTTTACTGCTCTTACATGATGTTCTCTGATGAGATGTC 2467
OY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlnGlyLysProTyrAsp 840
DB 2468 CATTTTGACATACCAAGTATATTTACTGAGTTTATAGAGAGGGCTGAGAAACCATATAT 2527
OY 841 LeuGlnIleTyrProGlnGlnArgHisSerIleArgValProGluSerGlyLysIleTyr 860
DB 2528 TTACAGATCTATCTCTGAGAGACACAGCATTAAGAGTTCTTGAATCGGAGAACTATAT 2587
OY 861 GlnLeuHisLeuLeuHisTyrLeuGlnGlnLeuAsnLeuGlySerArgIleAlaLeuLys 880
DB 2588 GAATCGATCTTTTGCACTACCTTCAAGAAACCTTGATCAGATATGCTGCTTAA 2647
OY 881 ValIle 882
DB 2648 GTGATA 2653

```

RESULT 3
ABK12892
ID ABK12892 standard; cDNA, 3106 BP.

AC ABK12892;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRTS-9 cDNA sequence.
XX
KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis; ss.
XX

```

OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 203..2851
FT /tag= a
FT /product= "Human protease PRTS-9"
PN WO200198468-A2.
PD 27-DEC-2001.
XX
XX
PF 13-JUN-2001; 2001WO-US19178.
XX
XX 16-JUN-2000; 2000US-212336P.
XX 22-JUN-2000; 2000US-213955P.
XX 29-JUN-2000; 2000US-215396P.
XX 07-JUL-2000; 2000US-216821P.
XX 14-JUL-2000; 2000US-218946P.
XX
XX (INCY- ) INCYTE GENOMICS INC.
XX
XX Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM,
XX Delgeane AM, Baughn MR, Nguyen DB, Lee EA, Hatalla A, Khan FA,
XX Walla NK, Yao WG, Lu DM, Patterson C, Tang YT, Walsh RT,
XX Azimzal Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L,
XX Kallick DA,
XX
XX WPI; 2002-090437/12.
XX P-PEDB; AAU74749.
XX
XX
XX Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
XX in the diagnosis, treatment and prevention of gastrointestinal (e.g.
XX gastritis), cardiovascular (e.g. atherosclerosis) and cell
XX proliferative (e.g. cancer) disorders -
XX
XX
XX Claim 5; Page 166-167; 177pp; English.
XX
XX
XX The present invention relates to twenty one new human proteases,
XX referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
XX polypeptides of the invention are useful in the diagnosis, treatment and
XX prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
XX Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
XX myocardial infarction, autoimmune/inflammatory e.g. acquired
XX immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
XX proliferative e.g. cancer, developmental e.g. Duchenne and Becker
XX muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
XX epilepsy and Alzheimer's disease and reproductive e.g. infertility and
XX endometriosis disorders. Numerous other examples of each disorder are
XX given in the specification. The present nucleic acid sequence encodes
XX the human protease PRTS-9 protein of the invention.
XX
XX
XX Sequence 3106 BP; 928 A; 633 C; 704 G; 841 T; 0 other;
XX
XX
XX
XX
XX Alignment Scores:
XX Pred. No.: 0
XX Score: 4700.00 Length: 3106
XX Percent Similarity: 100.00% Matches: 882
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 24 Gaps: 0
XX
XX
XX US-10-070-464-1 (1-882) x ABK12892 (1-3106)
XX
XX
XX 1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
XX
XX 203 ATGGACACACCAATGGAAACAGACAGCTGGTGTGATATTTGAAACCTGCGCAGACTGT 262
XX
XX 21 GlnGlnAsnIleGluSerGlnAspArgProLysLeuGlnProPheTyrValGlnArgTyr 40
XX
XX 263 GAGGAGATATTTGATACACGATCGCCCTTAATTTGAGGCTTTTATGTTGACGGGAT 322
XX
XX 41 SerTyrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
XX

```

323 TCCTGAGTCACTTAAAAAGCTGCTGCCATACAGAAAAATATCATGCTACATGATG 382
QY 61 AAlvSaIaProHiaSpHemcPheValIysArgAsnAspProAspGlyProHiaSer 80
DB 383 GCTAAGGCACCAATGATTTGATTTGGAAGAGAAATGATCCAGATGAGCCTCATCA 442
QY 81 AspArgIleTyrTyrIleuAlaMetSerGlyIuAsnArgGluAsnThrIleuPheTyrSer 100
DB 443 GACAGATCTTATTAATCTTCCATGCTGTGTGAGAAACAGAAAAATACCTGTTTATCT 502
QY 101 GJuIleProIysThrIleAsnArgAlaValIleuMetIleuSerTrpIysProIleu 120
DB 503 GAAATCCCAAAACATCATATAGAGACAGCTTAATGCTCTTGGAAAGCCTCTTTG 562
QY 122 AspIleuPheGlnAlaThrIleuAspTyrGlyMetIlyrSerArgGluGluIleuLeuArg 140
DB 563 GATCTTTTCAGCAACACTGACATAGATGATATCTCGAAGAAAGAACTATTAAGA 622
QY 141 GluArgIysArgIleGlyThrValGlyIleAsnIleAsnTyrAspTyrHisGlnGlySerGly 160
DB 623 GAAAGAAACGCAATGGAGACAGTCGGAATGCTTCTTACGATTAACCAAGAAAGTGA 682
QY 161 ThrPheIleuPheGlnAlaGlySerGlyIleTyrHisValIysAspGlyGlyProGlnGly 180
DB 683 ACATTTCTGTTTCAGCCGAGTGAATTTATCAAGTAAAGATGAGGCGCCACAGAGA 742
QY 181 PheThrGlnGlnProIleuArgProAsnIleuValGluThrSerCysProAsnIleArgMet 200
DB 743 TTTAGCAACACCTTAAAGCCCAATCTAGTGAACCTAGTTGCTCCACATACAGATG 802
QY 201 AspProIysIleuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
DB 803 GATCCAAATATATGCTGCTGATCCAGCTGATGCTGTTTATACATAGCAACATATTT 862
QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgIleuThrTyrValHisAsnGluLeu 240
DB 863 TGGATATCTTACATGCTACAGAGAGAGAAAGAGACTCACTATATGTCACATAGCTA 922
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValIleuGlnGlu 260
DB 923 GCCAATGAGAAAGAAATGCAAGATCAGCTGAGAGTCCGCTACTTGTCTCCAAAGAA 982
QY 261 PheAspArgTyrSerGlyTyrTrpTrpCysProIysAlaGluThrThrProSerGlyGly 280
DB 983 TTGTAGATATTTCTGGCTATTTGGTGTCTCCAAAGCTGAAACAACTCCAGTGTGT 1042
QY 281 LysIleLeuArgIleLeuTyrGluGluIuAsnAspGluSerGluValGluIleIleHisVal 300
DB 1043 AAATATCTTATGATATCTATATGAGAAATGATGATGATGAGTGAATATATCATGTT 1102
QY 301 ThrSerProMetIleuGluThrArgArgAlaAspSerPheArgTyrProIysThrGlyThr 320
DB 1103 ACATCCCTATGTTGAAACAGAGAGGAGGAGATCATCTCGTATCTTAAACACAGTACA 1162
QY 321 AlaAsnProIysValThrPheIysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
DB 1163 GCAATCTTAAAGTCACTTTTAAAGATGTCAGAAATATATGATGATGCTGAAAGAGATC 1222
QY 341 IleAspValIleAspIysGluLeuIleGlnProIleGluIleLeuPheGluGlyValGlu 360
DB 1223 ATAGAGTCAATAGATAGAACTAATTAACCTTTGAGATCTCTTTGAAAGAGATGAA 1282
QY 361 TyrIleAlaArgAlaGlyTrpThrProGluGlyIysTyrAlaTrpSerIleLeuLeuAsp 380
DB 1283 TATATGCGCAGAGCTGATGAGACTCTGAGGAGAAATATGCTTGGTCCATCTCATTAAT 1342
QY 381 ArgSerGlnThrArgIleuGlnIleValIleuIleSerProGluIleuPheIleProValGlu 400
DB 1343 CGCTCCCAAGCTCGCTACAGATAGTGTGATCTCACCGAATTTTATATCCAGATAGA 1402
QY 401 AspAspValMetGluIysGluArgIleuIleGluSerValProAspSerValThrProIleu 420
DB 1403 GATGATGTTATGAGAAAGCAGAGACTCATGAGTCACTGCTGATCTGTAGCGCACTA 1462

QY 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
DB 1463 ATATCTATGAAACAAACACAGCATCTGAAATATATCAAGACATCTTATGATTTT 1522
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIysThrGlyPhe 460
DB 1523 CCCCAAGTACAGAAAGAGAAATGAGTTATTTTGGTCTGCAATGCAAAACAGTTTC 1582
QY 461 ArgHisIleuTyrIysIleThrSerIleLeuIysGluSerIysTyrIysArgSerGly 480
DB 1583 CGTCAATTAACAAATATCATCTATTTAAAGAAAGCAAAATATTAACATCCAGTGTGT 1642
QY 481 GlyIleuProAlaProSerAspPheIysCysProIleIleGluGluIleAlaIleTrpSer 500
DB 1643 GAGCTGCTGCTCCAGATGATTTCAAGTGTCTTATCAAGAGAGATAGCAATTAACAGT 1702
QY 501 GlyIleuTrpGluValIleuGlyArgHisGlySerAsnIleGluValAspGluValArgArg 520
DB 1703 GGTGAATGAGAAAGTCTTGGCCGCAATGATCTTATATCCAGTTGATGAAAGTCAAGAG 1762
QY 521 LeuValTyrPheGluGluThrIlyAspSerProIleuGluHisIleIleuTyrValIleSer 540
DB 1763 CTGATATATTTTGAAGACCAAGACTCCCTTATGAGATCACTGATCGTATGATCAGT 1822
QY 541 TyrValAsnProGlyGluValThrArgIleuThrAspArgGlyTyrSerHisSerCysCys 560
DB 1823 TAGCTAATCTCGAAGAGTGAACAGGCTGATGACCGTGGCTATCCACATCTTCTGCTGC 1882
QY 561 IleSerGlnHisCysAspPhePheIleSerIlyrSerAsnGlnIysAsnProHisCys 580
DB 1883 ATCAGTCAAGACGTGCTCTTTATTAAGTAAATATATGATCAACAGAAATCAACTGT 1942
QY 581 ValSerIleuTyrIysIleuSerSerProGluIysAspAspProThrCysIysThrIysGluPhe 600
DB 1943 GTGTCCCTTACAAAGCATTAAGTCTGAAAGATGACCAACTGCAAAACAAAGAAATTT 2002
QY 601 TrpAlaThrIleLeuAspSerAlaGlyProIleuProAspTyrThrProProGluIlePhe 620
DB 2003 TGGGCCACCAATTTGATTCAGACAGTCCCTCTTCTGACTATATCTCTCCAGAAATTTTC 2062
QY 621 SerPheGluSerThrThrArgIlyPheThrIleuTyrGlyMetIleuTyrIysProHisAspIleu 640
DB 2063 TCTTTTAAAGTACTACTGATTTTACATTTGATGAGATGCTCTTCAAGGCTCATGATCTA 2122
QY 641 GlnProGlyIysIysTyrProThrValIleuPheIleTyrGlyGlyProGluValGluLeu 660
DB 2123 CAGCTCGAAGAAATATCTCTATCTGCTGCTCATATATGATGATGCTCTCAGAGTCACTG 2182
QY 661 ValAsnAsnArgPheIysGlyValIysTyrPheArgIleuAsnThrIleuAlaSerIleuGly 680
DB 2183 GTGATATATCGTTTAAAGAGTCAAGTATTTCCGCTGGAATACCTTACGCTCTCATGAT 2242
QY 681 TyrValValValIleAspAsnArgGlySerCysHisArgGlyIleuIysPheGluGly 700
DB 2243 TATGTGTGTATGATGATAGACACAGGGATCTCTGACCGAGGCTTAAATTTTGAAGGC 2302
QY 701 AlaPheIysTyrIysMetGlyGlnIleGluIleAspAspGlnValGluGlyIleuGlnTyr 720
DB 2303 GCCCTTAAATATTAAGAGGTCAAAATGAAGATCAAGTCAAGTGAAGAGATCCCAATAT 2362
QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740
DB 2363 CTAGCTCTCGATATGATTTTCACTTACATTCGATGATGATGATGATGATGATGATGAT 2422
QY 741 GlyGlyTyrIleuSerIleuMetAlaIleuMetGlnArgSerAspIlePheArgValAlaIle 760
DB 2423 GAGAGATACCTCTCCCGATGAGATTAATGCAAGATGATGATATCTTCAAGGATGATGAT 2482
QY 761 AlaGlyAlaProValThrIleuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
DB 2483 GCTGGGCCCCAGTCACTGTGTGATCTTATGATACAGATACAGAAACGTTATATG 2542

QY 761 GLYHISProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
 DB 2543 GGTACCTCTGACAGATGAACAGGAGCTATTACCTTACGATCTGGCCATGCAAGCAGAA 2602
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
 DB 2603 AAGTTCCTCTGACCAAAATGCTTACTGCTTACATGATGTTCTTGATGAGAAATGTC 2662
 QY 821 HISPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyLysProTyrAsp 840
 DB 2663 CATTTTGACATACACAGATATATTACTGAGTTTAAAGAGGCGCTGAGAAACCATATGAT 2722
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyLysIleTyr 860
 DB 2723 TTACAGATCTATCCCTCAGAGAGACACACAGCATTAAGATTCTGATGGGAGAACTATT 2782
 QY 861 GILeUHisIleLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLys 880
 DB 2783 GAACCTGACATCTTTTGCACCTACCTTCAGAAACCTTGATGACGATATGCTCTTAAA 2842
 QY 881 ValIle 882
 DB 2843 GTGATA 2848

RESULT 4
 AAC85694
 ID AAC85694 standard; cDNA; 3120 BP.
 XX AAC85694;

AC AAC85694;
 XX 29-JUN-2001 (first entry)

DE Nucleotide sequence of human DPP8.

XX Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
 KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
 KW growth hormone deficiency; glucose level; mucosal regeneration;
 KW non-insulin dependent diabetes mellitus; glucose intolerance;
 XX immunosuppression; ss.

OS Homo sapiens.

PH Key Location/Qualifiers
 FT CDS 214..2862
 FT /*tag= a
 FT /product= "Human DPP8"

PN WO200119866-A1.

PD 22-MAR-2001.

PF 11-SEP-2000; 2000MO-AU01085.

PR 10-SEP-1999; 99AU-0002762.
 PR 18-FEB-2000; 2000AU-0005709.

PA (UNSY) UNTV SYDNEY.

PI Abbott CA, Gorell MD;

DR WPI; 2001-281520/29.

DR P-PSDB; AAB47187.

PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
 PT substrates, identifying inhibitors of DPP8 catalytic activity which
 PT have therapeutic uses, and for detecting activated T cells
 PS Claim 16; Fig 2; 78pp; English.

CC This sequence encodes human dipeptidyl aminopeptidase (DPP8). DPP8
 CC has substrate specificity for H-Ala-Pro-PNA, H-Gly-Pro-PNA and
 CC H-Alg-Pro-PNA. Therefore, it is a prolyl oligopeptidase and a
 CC dipeptidyl peptidase, because it is capable of hydrolysing the

CC peptide bond C-terminal to Pro in each of these compounds. DPP8
 CC is homologous with human DPPIV. DPP8 is useful for cleaving a
 CC substrate, and for detecting an activated T cell which involves
 CC measuring the level of DPP8 gene expression in a T cell. The level
 CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
 CC in the cell. It is also useful for identifying a molecule capable
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules
 CC identified as inhibiting DPP8 catalytic activity may be useful for
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
 CC in non-insulin dependent diabetes mellitus and other disorders
 CC involving glucose intolerance, enhancing mucosal regeneration and
 CC as immunosuppressants.

CC Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

CC Alignment Scores:

CC Pred. No.: 0 Length: 3120
 CC Score: 4700.00 Matches: 882
 CC Percent Similarity: 100.00% Conservative: 0
 CC Best Local Similarity: 100.00% Mismatches: 0
 CC Query Match: 100.00% Indels: 0
 CC DB: 22 Gaps: 0

US-10-070-464-1 (1-882) x AAC85694 (1-3120)

QY 1 MetAlaAlaMetGluThrGlnGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
 DB 214 ATGGCAGCAGCATATGAAACAGACAGCTGGGCTTGAGATATTGAACTGGGACTG 273
 QY 21 GILeUHisIleLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLys 40
 DB 274 GAGAGAAATTTTAAATCAGACAGATCGCTTAAATTGAGACCTTTTATGTTGACCGTAT 333
 QY 41 SerTrpSerGlnLeuLysIleLeuAlaSerThrArgLysTyrHisGlyTyrMetMet 60
 DB 334 TCTCGAGTCAGCTTTAAAAGCTGCTTCCGACACAGAAATATATGCTCATAGATG 393
 QY 61 AlAlaAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
 DB 394 GCTAAGCAGCAGCATATGATTTGATTTGCAAGAGAAATGATCAGATGACCTCATTTCA 453
 QY 81 AspArgGluLeuTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
 DB 454 GACGAGATCTATTAATCTTCCATGCTGAGGAGAAACAGAAATACACTGTTTATTTCT 513
 QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
 DB 514 GAATTTCCCAAACTTCATATAGACAGACAGCTTTATGCTCTTGGAAAGCCTCTTTTG 573
 QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
 DB 574 GATCTTTTTCAGGCAACACTGACCTATGAGATGATTTCTCGAGAAAGAACTATTAAAG 633
 QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
 DB 634 GAAAGAAAAGCATTGGACAGTCGAAATGCTTTTACGATTTATACCAAGAGATGGA 693
 QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyLysProGlnGly 180
 DB 694 ACATTTCTGTTTCAACCGGTAAGTGAATTTATCAGCTTAAAGATGAGAGGCGCAAGAGA 753
 QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
 DB 754 TTTAACGAAACACTTTTAAAGCCCAATTCAGTGAAGAACTAGTTGTCCTCAACATACGAGATG 813
 QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
 DB 814 GATCCAAAATTAAGCCCGCTGATCCAGACTGGAATGCTTTTATACATAGCAACGATATT 873
 QY 221 TrpIleSerAsnIleValThrArgGluGluArgLeuThrTyrValHisAsnGluLeu 240
 DB 874 TGGATATCTTAAACATCTGTAACAGAGAAAGAAAGAGACTCATTTATGTGACATATGAGCTTA 933

2y	241	AlAsmMeTcLuGluNaSpAlaArgSerAglVValAlaThrPheValLeuGlnGlu	260
Db	934	GCCAACTGGAAAGAGATGCCAGATCACTGGAGTCCCTACTTTGTTCTCCAAAGAGA	993
Qy	261	PheAspArgTyLeuSerGlyTyTrpTrpCysProIysAgluThrThrProSerGlyGly	280
Db	994	TTTCATATGATTAATTCCTGGCTATTGGTGGTGTCCAAAGCTGAACCAACTCCAGGGTGT	1053
Qy	281	LysIleLeuArgIleLeuTyTyrGluGluAsnAspGluSerGluValGluIleIleHisVal	300
Db	1054	AAAATTCCTTGAATTCATATGAAAGAAATGATAACTGAGGGGAAATTAATTCAGTT	1113
Qy	301	ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyProIysThrGlyTyThr	320
Db	1114	ACATCCCTCATTTGTTGGAACCAAGAGGGCAGATTCAATTCCTTATCCTAAACAGGTACA	1173
Qy	321	AlaAsnProIysValThrPheLysMetSerGluIleMetIleAspAgluGlyArgIle	340
Db	1174	GCAATCCTTAAGTCACCTTTTAAGATGTCAGAAATAATGATTTGATGCTGAAGAGAGATC	1233
Qy	341	IleAspValIleAspLysGluLeuIleGlnProHegluIleAspPheGluGlyValGlu	360
Db	1234	ATGATGTCATAGAGTAAGAACTAATTCACCTTTTGAGATTCTATTGGAAGAGTTGAA	1293
Qy	361	TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp	380
Db	1294	TATATGCCAGAGCTGATGATGACTCTGAGGAAATATGCTGTGCTCATCTCTGAT	1353
Qy	381	ArgSerGluThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu	400
Db	1354	CGCTCCCAAGCTCGCTCAAGATATGTTGATCTCACTGAAATTAATTAATCCAGTGA	1413
Qy	401	AspAspValMetGluArgGluArgLeuIleGluSerValProAspSerValThrProLeu	420
Db	1414	GATATGTTATGGAAGAAAGCAGAGACTCATTTGATGATGAGCTGATTCGTAGCCACTA	1473
Qy	421	IleIleTyTyrGluGluThrThrAspIleLeuTrpIleAsnIleHisAspIlePheHisValPhe	440
Db	1474	ATTATCTTGAAGAAACAACAAGACATCTGGATAATATCCAGATCTTTCATAGTTT	1533
Qy	441	ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeuThrGlyPhe	460
Db	1534	CCCCAAAGTCAGAGAGAGAAATTGATTTATTTTGGCTGTGATGCAAAACAGGTTTC	1593
Qy	461	ArgHisLeuTyTyrIleThrSerIleLeuLysGluSerLysTyTyrLeuArgSerGly	480
Db	1594	CGTCAATTATACAAATATACATCTATTTTAAAGAAACAATATTAACATCCAGTGGT	1653
Qy	481	GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer	500
Db	1654	GGGCGCGCTGCTCAAGATGATTTCAAGTCTCTATCAAAAGAGAAATACGATTTCCAGT	1713
Qy	501	GlyGluTrpGluValLeuGlyValArgHisGlySerAsnIleGlnValAspGluValArgArg	520
Db	1714	GGTGATGGGAAGTTCTTGCGCGGATGGATTAATATCAAGTTGATGAAGTCAGAAAG	1773
Qy	521	LeuValTyTyrPheGluGlyThrTyAspSerProLeuGluHisHisLeuTyTyrValaLys	540
Db	1774	CTGGATATTTTGAAGGCAACCAAGACTCCCTTTAGAGCATCACTGTACATGACGT	1833
Qy	541	TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyTyrSerHisSerCysCys	560
Db	1834	TAAGTAATCCCTGGAGAGGTGACAAAGGCTGACGACCGGTACTCATCAATCTTGCTGC	1893
Qy	561	IleSerGlnHisCysAspPhePheIleSerLysTyTyrSerAsnGlnLysAsnProHisCys	580
Db	1894	ATCAGTCACACCTGACACTCTTTTAAAGATAGTATAGTAACCAAGAGATCCACAGT	1953
Qy	581	ValSerLeuTyTyrLysLeuSerSerProGluAspAspProThrCysLysThrTyGluPhe	600
Db	1954	GTGTCCCTTTACAGATCAAGTCAAGTCTCGAAGATGACCAACTTGCAAAACAAAGGATTT	2013
Qy	601	TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyThrProProGluIlePhe	620

	Accession	Gene	Protein	Length (aa)
Db	2014	TGGGACCACTTTGGATTACGACAGGTCCTCTTCTGACTATACCTCTCCGAAATTTTC		2073
Oy	621	SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIysProHisAspLeu		640
Db	2074	TCTTTTGAAGTACTACTGCGATTACATGTGATGGATGCTCTCAAGCCCTCATGATCTA		2133
Oy	641	GlnProGlyIysIysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu		660
Db	2134	CAGCCTGGAAAGAAATATCTCTACTGTCCTGTCTCATATATGGTGGTCTCCAGGTCCAGTTG		2195
Oy	661	ValAsnAsnArgPheIysGlyValIleSerTyrPheArgLeuAsnThrLeuAlaSerLeuGly		680
Db	2194	GTGAATATATGCGTTTAAAGAGTCAGATATTTCCGCTTGAAATACCTTGCTCTAGGT		2255
Oy	681	TyrValValValIleAspAsnArgIysSerCysHisArgGlyLeuIysPheGluGly		700
Db	2250	TATGTGTGTGTAAGATGATGACACAGGGAGTCCTGTACCCAGGGCTTAAATTTGAAGGC		2314
Oy	701	AlaPheIysTyrIysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuTyr		720
Db	2314	GCTTTTAAATATTAATATGGCTCAATATGAAATTGACGATCCAGTCCGGTGAAGACTCCAAAT		2377
Oy	721	LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr		740
Db	2374	CTACCTCTTCATATGATTTATTTATTTAGACTTAAATGCTGTGGCATTCACAGGCTGTCCTAT		2433
Oy	741	GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle		760
Db	2434	GGAGGATACCTCTCTCCCTGATGGCATTTATGACAGAGTCAGATATCTTCAGGGTTCTAAT		2493
Oy	761	AlaGlyAlaProValThrLeuTyrPheTyrAspThrGlyTyrThrGlnArgTyrMet		780
Db	2494	GCTGGGGCCCAAGTCACCTGTGGATCTTCTATATGATACAGATACAGGAACGGATATATAG		2555
Oy	781	GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu		800
Db	2554	GATCACCTCGACACAGATGATACAGGGCTATTTACTTATGGATCTGTGGCCATTCAGACAGA		2611
Oy	801	LysPheProSerGlnProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal		820
Db	2614	AAATTCCCTCTCGAACCAAAATGGTTTACTGCTTACATGATGGTTTCTGTGATGAGAAATGTC		2677
Oy	821	HisPheAlaHisSerThrSerIleLeuLeuSerPheLeuValArgAlaGlyIysProTyrAsp		840
Db	2674	CATTTTGCACATACACAGATATATTAATGATGTTTTTATGAGAGGGCTGGAAAGCCATATGAT		2733
Oy	841	LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr		860
Db	2734	TTACAGATCTTATCTTCAGAGAGACACAGCATTAAGAGTTTCTGAAATCGGAGAACCTTAT		2793
Oy	861	GluLeuHisIleLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuIys		880
Db	2794	GAACTGCACATCTTTGCACACTACTTCAAGAAAACTTGGATCAGTATGTGCTGTATAAA		2855
Oy	881	ValIleIle 882		
Db	2854	GTGATA 2859		
RESULT 5				
ID	AAD38956	standard; cDNA, 3120 BP.		
AC	AAD38956;			
XX				
DT	23-SEP-2002	(first entry)		
XX				
DE		Human dipeptidyl peptidase 8 (DPP8) cDNA.		
XX				
KM		Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;		
KM		autoimmunity; human immunodeficiency virus; HIV infection; cytostatic;		
KM		graft rejection; antidiabetic; antiinflammatory; immunosuppressive;		
KM		antiviral; enzyme; gene; ss.		

```

XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH CDS 214..2862
FT /*tag= a
FT /product= "Human DPP8 protein"
XX
XX MO200234900-A1.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001W0-AU01388.
XX
XX 27-OCT-2000; 2000AU-0001078.
XX
XX (UNSY ) UNIV SYDNEY.
XX
XX Abbot CA, Gorrell MD;
XX
XX WPI: 2002-454646/48.
XX
XX P-PSDB; AAE24170.
XX
XX New dipeptidyl peptidase (DPP) peptides, useful for screening
XX inhibitors of DPP catalytic activity, which may be employed to treat
XX e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV infection -
XX
XX Example; Fig 1; 91pp; English.
XX
XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX polynucleotides encoding such proteins. The DPP peptides are useful for
XX screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV (human immunodeficiency virus) infection. The present
XX sequence is human DPP8 cDNA.
XX
SQ Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 3120
Score: 4700.00 Matches: 882
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0

US-10-070-464-1 (1-882) x AAD38956 (1-3120)

QY 1 MetaAlaAlaMetGluThrGluGluLeuGlyValGluIlePheGluThrAlaAspCys 20
Db 214 ATGCACAGCAGATGGAACAGACAGCTGGGTGTGAGATATTTGAACCTGCGACATGT 273
QY 21 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40
Db 274 GAGGAGATATTTGATGACAGATCGGCTTAATTTGAGGCTTTTATGTTGAGCGGAT 333
QY 41 SerThrSerGluLeuGlyGlyLeuLeuAlaAspThrArgGlyTyrHisGlyTyrMetMet 60
Db 334 TCCGGAGTCAAGCTTAAAGCTCTGCTCCCATACCGAATAATATCATGCTACATGATG 393
QY 61 AlaAlaAlaProHisAspPheMetPheValIleArgGluGluGluGluGluGluGluGlu 80
Db 394 GCTAAGCACCACATGATTTTCATGTTGTGAGAGATGATCCAGATGAGACCTCATTC 453
QY 81 AspArgGlyLeuTyrTyrLeuAlaMetSerGlyGluGluGluGluGluGluGluGluGlu 100
Db 454 GACAGATCTATTAACCTTGCATGCTGCTGAGAGACAGAGAAATATACAGTTTATCT 513
QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeuLeu 120
Db 514 GAAATTCCTCAAACTATCAATAGAGAGAGCTTAAATGCTCTCTTGGAGCCCTTTTG 573

```

```

QY 121 AspLeuPheGluAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
Db 574 GATCTTTTTCAGGCAACACTGACATGATGATGATTTCTCGAAGAGAACTATTAAG 633
QY 141 GluArgGlyArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlyGlySerGly 160
Db 634 GAAAGAAACGATTTGAAACAGTCCGAATTTGCTTTTACGATTAATCCACAGAACTGCA 693
QY 161 ThrPheLeuPheGluAlaGlySerGlyIleTyrHisValIleAspGlyGlyProGlnGly 180
Db 694 ACATTTCTGTTCCAGCCGGTGTGGAATTTATACGTAAGAAATGAGAGGCCACAGGA 753
QY 181 PheThrGlnGluProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db 754 TTTACGCAACACACTTTAAGCCCATCATGAGAAACTAGTTGCTCCAAACATACGATG 813
QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
Db 814 GATCCAAATTAATGCCCCGCTGATCCAGCTGATTTGCTTTATACATGACACGATATT 873
QY 221 TrpIleSerAsnIleValThrArgGluGluGluArgGlyLeuThrTyrValHisAsnGluLeu 240
Db 874 TGGATATCTAATCATGTAACGACAGAGAAAGAGACTCATTAATGTGCAAAATGAGCTA 933
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
Db 934 GCCAACATGGAAGAGATGCCAGATCAGCTGAGTCCGTAACCTTTGTTCTCCAGAGAGA 993
QY 261 PheAspArgTyrSerGlyTyrTyrTrpCysProLysAlaGluThrThrProSerGlyGly 280
Db 994 TTGTGATATATTCGGCTATTTGGTGGTGTCCAAAGCTGAAACACTCCCGATGTGCT 1053
QY 281 LysIleLeuAlaGlyIleLeuTyrGluGluGluGluGluGluGluGluGluGluGluGlu 300
Db 1054 AAAATTTCTTAATTTCTATGAGAAATATGATGATCTGAGTGAATTAATTAATGATGTT 1113
QY 301 ThrSerProMetLeuGluThrThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db 1114 ACATCCCTTATGTTGGAACAAAGAGGCGACAGTTCAATCCGTTATCCTTAACAGATACA 1173
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db 1174 GCAATCTTAAGTCACTTTTAAGATGTCAGAAATAAAGATTTGATGCTGCAAGAGAGATC 1233
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db 1234 ATAGATGCTAATGATTAAGAACTAATTTCACTTTTGATTTGATTTGAAGAGAGTTGAA 1293
QY 361 TyrIleAlaArgAlaGlyTyrTrpThrProGluGlyLysTyrAlaTyrSerIleLeuLeuAsp 380
Db 1294 TATATTGCGAAGCTGATGATGATCTCTGAGGAAATATGCTTGTCATCTACTAAT 1353
QY 381 ArgSerGlnThrArgLeuGluIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db 1354 CGTCCCAAGATCCCTTACATATGATGTGATTCACCTGATTAATTAATCCAGTAGAA 1413
QY 401 AspAspValMetGluArgGluArgLeuIleGluSerValProAspSerValThrProLeu 420
Db 1414 GATGATTTATGAGAAAGCAGAGACTCATTTAGTCAAGTCTGATTTCTGTAAGCCACTA 1473
QY 421 IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
Db 1474 ATTATCTATGAGAAACAAAGACATCTGATTAATATTCATGATCATCTTTCACTTTT 1533
QY 441 ProGlnSerHisGluGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
Db 1534 CCCCAGAGTCAAGAGAGAAATTTAATTTATTTTCCCTGATCAACAAACAGGTTTC 1593
QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerTyrLysArgSerSerGly 480
Db 1594 CGTCAATTAACAAATTAATCATATTTTAAAGAAAGCAAAATTAACGATTCACAGTGGT 1653
QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500

```